



Description of a global marine particulate organic carbon-13 isotope data set

Maria-Theresia Verwega^{1,2}, Christopher J. Somes¹, Markus Schartau¹, Robyn Elizabeth Tuerena³, Anne Lorrain⁴, Andreas Oschlies¹, and Thomas Slawig²

¹GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

²Department of Computer Science, Kiel University, Kiel, Germany

³Scottish Association for Marine Science, Dunstaffnage, Oban, PA37 1QA, UK

⁴LEMAR, Univ Brest, CNRS, IRD, Ifremer, 29280 Plouzané, France

Correspondence: Christopher J. Somes (csmes@geomar.de)

Received: 14 May 2021 – Discussion started: 26 May 2021

Revised: 28 September 2021 – Accepted: 29 September 2021 – Published: 26 October 2021

Abstract. Marine particulate organic carbon stable isotope ratios ($\delta^{13}\text{C}_{\text{POC}}$) provide insights into understanding carbon cycling through the atmosphere, ocean and biosphere. They have for example been used to trace the input of anthropogenic carbon in the marine ecosystem due to the distinct isotopically light signature of anthropogenic emissions. However, $\delta^{13}\text{C}_{\text{POC}}$ is also significantly altered during photosynthesis by phytoplankton, which complicates its interpretation. For such purposes, robust spatio-temporal coverage of $\delta^{13}\text{C}_{\text{POC}}$ observations is essential. We collected all such available data sets and merged and homogenized them to provide the largest available marine $\delta^{13}\text{C}_{\text{POC}}$ data set (<https://doi.org/10.1594/PANGAEA.929931>; Verwega et al., 2021). The data set consists of 4732 data points covering all major ocean basins beginning in the 1960s. We describe the compiled raw data, compare different observational methods, and provide key insights in the temporal and spatial distribution that is consistent with previously observed large-scale patterns. The main different sample collection methods (bottle, intake, net, trap) are generally consistent with each other when comparing within regions. An analysis of 1990s median $\delta^{13}\text{C}_{\text{POC}}$ values in a meridional section across the best-covered Atlantic Ocean shows relatively high values ($\geq -22\text{\textperthousand}$) in the low latitudes ($< 30^\circ$) trending towards lower values in the Arctic Ocean ($\sim -24\text{\textperthousand}$) and Southern Ocean ($\leq -28\text{\textperthousand}$). The temporal trend since the 1960s shows a decrease in the median $\delta^{13}\text{C}_{\text{POC}}$ by more than 3\textperthousand in all basins except for the Southern Ocean, which shows a weaker trend but contains relatively poor multi-decadal coverage.

1 Introduction

Carbon is an essential element for life, and it regulates climate via its atmospheric form CO₂, a long-living greenhouse gas. Understanding carbon cycling is fundamental to reliably projecting changes in the Earth's future climate. Carbon is subject to transformation and cycling throughout the ocean, land and atmosphere. It is a major part of organic matter of all living organisms which can both consume (e.g., photosynthesis) and produce (e.g., respiration) inorganic carbon. Besides the natural cycling processes, the total amount and distribution of carbon is strongly perturbed by human activity caused by industrialization, most notably due to fossil fuel emis-

sions, deforestation, farming, cement production and other industrial processes. Anthropogenic CO₂ emissions are one of the main driving forces of modern climate change which is likely to continue in the future (IPCC, 2013). Only about 60 % of anthropogenic CO₂ emissions have been compensated for by natural sinks, including the dissolution of inorganic carbon in the ocean. This means the atmosphere has already been enriched with anthropogenic carbon by about 880 Gt CO₂ since 1750 (IPCC, 2014), which is driving the increase in global temperature levels. The ocean serves as an important buffer as it absorbs a significant amount of an-

thropogenic carbon, with the ocean interior being the largest readily exchangeable reservoir of carbon in the Earth system.

Marine phytoplankton convert dissolved inorganic carbon (e.g., aqueous CO₂) into organic carbon via photosynthesis in the euphotic surface layer. This organic carbon forms the base of the food web for higher trophic levels in marine ecosystems. Some particulate organic carbon (POC) sinks down to ocean depths, where it either is respiration back to dissolved inorganic carbon by heterotrophic organisms or becomes buried in ocean sediments (Suess, 1980). This process is known as the soft-tissue biological carbon pump, an important mechanism for sequestering carbon to the deep ocean from the atmosphere (Volk and Hoffert, 1985; Banse, 1990; McConaughey and McRoy, 1979). Since the deep ocean has a residence time of about a millennium, it is a key carbon reservoir influencing long-term climate change.

Carbon isotopes provide additional insights into the cycling of carbon in the Earth system (Zeebe and Wolf-Gladrow, 2001). The element carbon exists in two naturally occurring stable isotopes, ¹²C and ¹³C, with abundances of around 98.9 % and 1.1 %, respectively. Knowledge of their pathways through carbon reservoirs can support deeper understanding of carbon transfer and can help identify carbon sources with different isotopic ratios (Rounick and Winterbourn, 1986). Relative abundances of carbon isotopes are usually given in δ notation, which is based on the carbon isotope ratio $\frac{^{13}\text{C}}{^{12}\text{C}}$, standardized and given in parts per thousands as

$$\delta^{13}\text{C} = \left(\frac{\frac{^{13}\text{C}}{^{12}\text{C}}}{R_{\text{std}}} - 1 \right). \quad (1)$$

The constant $R_{\text{std}} = 0.0112372$ is a standard ratio, originally referring to the calcareous fossil Pee Dee Belemnite. The values ¹²C and ¹³C are the absolute concentrations of the individual isotopes (Hayes, 2004).

Distributed within the carbon cycle, the fractionation of $\delta^{13}\text{C}$ is influenced by biological and thermodynamic processes (Gruber et al., 1999). Air-sea gas exchange plays a dominant role at the ocean surface. Phytoplankton photosynthesis and POC remineralization increase their influence in the ocean interior (Gruber et al., 1999; Morée et al., 2018). The processes are dependent on circulation and temperature, and thus their individual influence varies with geographic location (Gruber et al., 1999; Schmittner et al., 2013).

Phytoplankton preferentially incorporate (i.e., fractionate) the lighter ¹²C carbon isotope into its organic matter. This fractionation causes phytoplankton organic $\delta^{13}\text{C}$ to be 10‰ to 25‰ lower than that of inorganic $\delta^{13}\text{C}$, which depends on a variety of environmental, ecological and physiological conditions (e.g., Popp et al., 1989, 1998; Rau et al., 1989, 1996). The main factors that control phytoplankton fractionation are concentrations of CO₂(aq), species-specific effects enforced by the phytoplankton composition and the cellular growth rate, although uncertainties remain regarding the quantifica-

tion of the specific processes and mechanisms that cause variations in phytoplankton fractionation (e.g., Fry, 1996; Laws et al., 1995; Popp et al., 1998; Bidigare et al., 1997; Cassar et al., 2006).

$\delta^{13}\text{C}_{\text{POC}}$ provides insights into physical and biological carbon cycle processes in the ocean (e.g., Fry and Sherr, 1989). It helps to diagnose carbon pathways from the atmosphere to the deep ocean including the biological carbon pump (e.g., Jasper and Hayes, 1990; Popp et al., 1989; Freeman and Hayes, 1992) and assists reconstruction of oceanic carbon cycling and even plankton cell size and community structure (e.g., Tuerena et al., 2019; Lorrain et al., 2020). For example, anthropogenic carbon emissions have a distinctly low $\delta^{13}\text{C}$ content, making $\delta^{13}\text{C}$ a useful property for tracing anthropogenic carbon throughout the Earth system (Eide et al., 2017; Levin et al., 1989; Ndeye et al., 2017). Atmospheric $\delta^{13}\text{C}_{\text{CO}_2}$ has decreased from $-6.5\text{\textperthousand}$ in preindustrial times to $-8.4\text{\textperthousand}$ presently (Rubino et al., 2013). The measurable decrease due to anthropogenic fossil carbon emissions is known as the Suess effect (Keeling, 1979), which enters the ocean via air-sea gas exchange. However, changes in marine $\delta^{13}\text{C}_{\text{POC}}$ are also significantly influenced by changes in phytoplankton fractionation due to other anthropogenic controls. For example increasing CO₂(aq) concentrations increase surface $\delta^{13}\text{C}$ fractionation (Young et al., 2013) and changes in phytoplankton composition and temperature influence phytoplankton growth rates and $\delta^{13}\text{C}$ fractionation over the air-sea interface (Zhang et al., 1995). But determination of the driving processes(es) of $\delta^{13}\text{C}_{\text{POC}}$ spatial and temporal trends remains a challenge. We also stress that all of these processes are sensitive to temperature changes which adds additional complexity to understanding how fractionation may change in space and time. A better understanding of the contributions from all of these effects requires a robust global data set of $\delta^{13}\text{C}_{\text{POC}}$.

Theoretical projection and understanding of changes associated with $\delta^{13}\text{C}_{\text{POC}}$ can be executed by models of different scales, which include $\delta^{13}\text{C}_{\text{POC}}$ circulation. Earth system models serve to simulate and test hypotheses in different scenarios as unbiased assessments (e.g., IPCC, 2014) and may support future decision-making. Besides resolving the mass flux of carbon, many models also simulate stable carbon isotopes (e.g., Schmittner and Somes, 2016; Buchanan et al., 2019; Hofmann et al., 2000; Jahn et al., 2015; Tagliabue and Bopp, 2008; Morée et al., 2018; Magozzi et al., 2017). For reliable calibrations and validations of such process-based mechanistic models, a spatially and temporally comprehensive data set is essential. This additional constraint provided by marine $\delta^{13}\text{C}_{\text{POC}}$ assists the reconstruction of oceanic carbon cycling including how much anthropogenic carbon is entering marine ecosystems and being exported to the deep ocean. But until today, there has been a lack of suitable data sets as constraints. This results in large and mostly unknown uncertainties in model results.

Data sets of marine $\delta^{13}\text{C}_{\text{POC}}$ improve our understanding of marine carbon cycling by providing another independent constraint. Recent model approaches support long-term past climate projections (Tjiputra et al., 2020) and assess estimations of the Suess effect (Liu et al., 2021). To date, numerous individual $\delta^{13}\text{C}_{\text{POC}}$ data sets exist, while the number of accessible, merged data sets is lacking. Existing merged data sets contain data from several sources but have often been focused on a specific region or process (e.g., Goericke, 1994; Tuerena et al., 2019). Individual data sets are usually collected during a specific cruise or time series station and are often neglected since they contain relatively few data. Such data sets can easily be accessed on data platforms such as PANGAEA and, when combined, they can represent an important and significant source of data.

In this study, we provide a novel merged seawater $\delta^{13}\text{C}_{\text{POC}}$ data product (Verwega et al., 2021) that – to our knowledge – contains the most expansive spatio-temporal coverage to date. It contains all available $\delta^{13}\text{C}_{\text{POC}}$ seawater data from PANGAEA and the merged data sets by Goericke (1994), Tuerena et al. (2019) and Young et al. (2013), as well as unpublished data from different cruises by Anne Lorrain. No data were excluded, even if sampled at extreme locations (e.g., trenches, hydrothermal vents). The metadata comprise information about the sampling location, time, depth and method as well as the original source, which makes original raw data values, methods, and further technical description easily accessible. Provided data files are Network Common Data Form (NetCDF) files interpolated onto two different global grids and a csv file that includes the data and their anomalies with respect to their overall mean together with all corresponding available meta-information.

The paper is structured as follows: we provide a brief overview of $\delta^{13}\text{C}_{\text{POC}}$ data acquisition in Sect. 2 and their compilation and metadata in Sect. 3. The characteristics of the collected $\delta^{13}\text{C}_{\text{POC}}$ data are shown in Sect. 4. We present their spatial distribution in Sect. 5 and temporal distribution in Sect. 6. Lastly, we provide a short summary and concluding remarks.

2 Data acquisition

The data set includes 4732 entries for $\delta^{13}\text{C}_{\text{POC}}$ from 185 different sources and ranges from the 1960s to the 2010s. In addition to many data sets from the data platform PANGAEA, we included unpublished data provided by Anne Lorrain and the data products from Tuerena et al. (2019), Goericke (1994) and Young et al. (2013). The adjustments that we conducted are described in the following.

2.1 Data sources

As a basis of our data set, we chose the 1990s data collection by Goericke (1994). This was established to investigate variations in $\delta^{13}\text{C}_{\text{POC}}$ with temperature and latitude.

The $\delta^{13}\text{C}_{\text{POC}}$ sample data and measurements were conducted by investigating zooplankton, net plankton or particulate organic matter. We cross-checked and extended this data set by looking up all available primary sources. Goericke (1994) originally included 476 $\delta^{13}\text{C}_{\text{POC}}$ data points from 17 contributions. The largest contributions came from Fischer (1989) with 107 entries, Fontugne et al. (1991) with 97, and Fontugne and Duplessy (1981, 1978) with 78. Large extensions were possible, e.g., in the Fischer (1989) and Eadie and Jeffrey (1973) data sets, incorporating more than 70 additional data points from these primary sources. With this extension, we could increase the data set to 626 data points for $\delta^{13}\text{C}_{\text{POC}}$.

We collected most data from the PANGAEA data platform, an open-access online library archiving and providing geo-referenced Earth system data, hosted and monitored by the Alfred-Wegener-Institut (2020) – Helmholtz Centre for Polar and Marine Research (AWI) – and the Center for Marine Environmental Sciences, University of Bremen (MARUM). With the data made available therein, we could further extend the data set by an additional ~ 3500 measurements of $\delta^{13}\text{C}_{\text{POC}}$. Most $\delta^{13}\text{C}_{\text{POC}}$ data from PANGAEA are associated with samples collected during the Joint Global Ocean Flux Study (JGOFS, 2020), with more than 2000 $\delta^{13}\text{C}_{\text{POC}}$ data points. Additionally, 529 samples are contributions by the Antarctic Environment and Southern Ocean Process Study (AESOPS, 2020), 342 are by the Archive of Ocean Data (EuroBIS Data Management Team, 2020) and 279 are by the SFB313 research project (Thiede et al., 1988).

Other collected data were provided by Robyn Tuerena and Anne Lorrain. Robyn Tuerena provided a data contribution coming from the data set mentioned in Tuerena et al. (2019), which we will refer to as the Tuerena data set. This contains 595 data points including 501 from Young et al. (2013) and covers samples within the euphotic zone and an observation time frame of 1964–2012. Moreover, we included 69 unpublished data points provided by Anne Lorrain, covering the years 2012–2015 and sampled during the cruises CASSIOPEE, PANDORA, OUTPACE, NECTALIS-3, NECTALIS-4 and KH-13. We refer to this data set as the Lorrain data set.

A recent collection of 303 measurements of $\delta^{13}\text{C}_{\text{POC}}$ has been provided by Close and Henderson (2020), largely based on data gathered from individual publications referenced therein. Since our analyses originally relied on data sources that differed from those of Close and Henderson (2020), we find our collection to be as yet incomplete. Especially measurements from national databases might provide a huge future benefit.

2.2 Adjustments made

All data were taken with as many details as possible from the sources and have been reshaped to fit the structure of the data set. No rounding or cutoff of detailed data was

made. Spatial coordinates originally given as depth intervals were replaced by their respective midpoints. Time intervals were not changed in this way. If they contained just 1 month or year, this was included; otherwise the time information was omitted. Sample depth given as “surface” was denoted as 1 m. Longitude values were converted to the format $[-180^\circ, 180^\circ]$ by the transformation

$$\text{Long}_{\text{new}} = \begin{cases} \text{Long}_{\text{old}} - 360^\circ & \text{for all } \text{Long}_{\text{old}} \in (180^\circ, 360^\circ] \\ \text{Long}_{\text{old}} & \text{otherwise.} \end{cases} \quad (2)$$

Wherever possible the data were taken from their original publication. Changes made to the data by Goericke (1994) are described in Table 1 and changes to all other data in Table 2. The complete structure is presented in Table 3.

Most data listed in the Goericke (1994) data set could be gathered from the original publications directly. Some data are not accessible from an original source, including those data labeled as “Harrison”, “Hobson” and “Schell”, which were included as unpublished data by personal communication in Goericke (1994). Also, we could not identify the original data sources of “Voss (1991)” and “Sackett et al. (1966)”. Data from these sources are used as provided by Goericke (1994). All other data could be directly compared with and linked to their origin. According to Table 3 we complemented the data with the month, year, depth, sample method, cruise, trap duration and references wherever available. Special notes given in Goericke (1994) were conserved in our “project/cruise”-named meta-information. Rounded values were adjusted to their source values as well as data with interchanged longitudinal information, which is shown in detail in Table 1.

In two cases we identified multiple $\delta^{13}\text{C}_{\text{POC}}$ data sets from a single event (time, place, investigator) where the data had been subject to different stages of processing or different types of measurement: in Westerhausen and Sarnthein (2003), we chose the “mass spectrometer” data set because this was the originally measured one. In Trull and Armand (2013a, b), we used the “blank corrections” data set of $\delta^{13}\text{C}$, since this set of $\delta^{13}\text{C}_{\text{org}}$ values is recommended to be considered (Trull and Armand, 2001).

The primary source of the Tuerena and Lorrain data was mentioned in our data set in the project/cruise column. In the data set from Tuerena et al. (2019), this was originally labeled as “source” and in the Lorrain data set as “campaign”. In both data sets the longitude was converted to $[-180^\circ, 180^\circ]$ from a $[0^\circ, 360^\circ]$ format by Eq. (2). In the data of MacKenzie et al. (2019) we deleted a typo where the depth value was set equal to the negative longitude value. We disregarded the trap duration given in Voss and von Bodungen (2003), which was given as the negative value -1 .

3 Content and structure of the data set

The data collection is made available in files of raw and interpolated values (Verwega et al., 2021). The raw data are

in a csv file that includes the $\delta^{13}\text{C}_{\text{POC}}$ measurements, their anomalies with respect to their mean and all available meta-information. The interpolated data are provided as NetCDF files on two different global grids: a $1.8^\circ \times 3.6^\circ$ resolution and 19 depth layers from a model that simulates $\delta^{13}\text{C}_{\text{POC}}$ (e.g., Schmittner and Somes, 2016), in the following referred to as the UVic grid, and the $1^\circ \times 1^\circ$ resolution and 102-depth-layer grid of the World Ocean Atlas (Garcia et al., 2018), in the following referred to as the WOA grid. Interpolation required the availability of the full spatial information (latitude, longitude and depth) of included $\delta^{13}\text{C}_{\text{POC}}$ data to locate them on the grid.

On the WOA grid we provide 13 NetCDF files containing only data with full spatio-temporal metadata: one averages all observations from each year together, each year accounting for a time increment on the time axis, and the other 12 files average only observations from an individual month with again each year accounting for a time increment on the time axis. These files provide a variety of analysis opportunities but also limited the content of $\delta^{13}\text{C}_{\text{POC}}$ data.

On the UVic grid we provide seven individual NetCDF files: six of them each represent one of the decades from the 1960s to the 2010s containing all data which were able to be assigned to their respective decade. One file contains all available $\delta^{13}\text{C}_{\text{POC}}$ data completely independent of their measurement time. This individual provision of data on a decadal and overall timescale increases the fraction of usable $\delta^{13}\text{C}_{\text{POC}}$ data for the following analyses.

3.1 Raw data file

The csv-format data file includes $\delta^{13}\text{C}_{\text{POC}}$ measurements, anomalies and meta-information in its columns. A full description of the content, value range and coverage of the individual columns is given in Table 3. Anomalies of $\delta^{13}\text{C}_{\text{POC}}$ were calculated, based on the arithmetic mean of the full data collection. The mean was calculated, rounded to two digits after the floating point and used as

$$\text{mean}_{\delta^{13}\text{C}_{\text{POC}}} = -23.96\%. \quad (3)$$

Anomalies contain all the relevant information with respect to the variability in the $\delta^{13}\text{C}_{\text{POC}}$ data in space and time. This way it becomes easier to analyze bias information separately, e.g., during the first steps of model calibration.

The reference includes the citations in as much detail as possible. Wherever available, this is taken from the original source. Otherwise, we tried to include the author, title, publication year and platform, and DOI. For unpublished data like Harrison’s (unpublished data, quoted from Goericke, 1994) from the Goericke (1994) data set or those included by the co-authors, we specified from where we took the data.

Coordinates are given in decimal degrees over $[-90^\circ, 90^\circ] \times [-180^\circ, 180^\circ]$. The sample depth is given in meters measured positively from the ocean surface downwards. Data published as measured at 0 m were included

Table 1. Changes that were introduced into data taken from Goericke (1994): the first column names the publication or author of the primary data set. The second column lists in which part of the data we applied changes. The third and fourth columns show what these changes were, and the last column gives the reason for this.

Data set	Changed	From	To	Reason
Degens et al. (1968)	long	Goericke (1994)	source value	E and W interchanged
Eadie and Jeffrey (1973)	long	Goericke (1994)	source value	E and W interchanged
Fischer (1989)	long	Goericke (1994)	source value	E and W interchanged
Fontugne and Duplessy (1978)	long	Goericke (1994)	source value	E and W interchanged
Fontugne and Duplessy (1981), MD13 Osiris III	long	Goericke (1994)	source value	E and W interchanged
Francois et al. (1993)	long	Goericke (1994)	source value	E and W interchanged
Harrison*	long	Goericke (1994)	source value	E and W interchanged
Sacket et al. (1965)	long	Goericke (1994)	source value	E and W interchanged
Saupe et al. (1989)	long	Goericke (1994)	source value	E and W interchanged
Wada et al. (1987)	long	Goericke (1994)	source value	E and W interchanged
Eadie and Jeffrey (1973)	lat, long	Goericke (1994)	source value	rounded in Goericke (1994)
Fischer (1989) all but INDOMED leg 12	lat, long	Goericke (1994)	source value	rounded in Goericke (1994)
Fontugne and Duplessy (1978)	lat, long	Goericke (1994)	source value	rounded in Goericke (1994)
Fontugne and Duplessy (1981)	lat, long	Goericke (1994)	source value	rounded in Goericke (1994)
Francois et al. (1993)	lat, long	Goericke (1994)	source value	rounded in Goericke (1994)
Sacket et al. (1965)	lat, long	Goericke (1994)	source value	rounded in Goericke (1994)
Eadie and Jeffrey (1973)	$\delta^{13}\text{C}_{\text{POC}}$	not included	added	not included in Goericke (1994)
Fischer (1989)	$\delta^{13}\text{C}_{\text{POC}}$	not included	added	not included in Goericke (1994)
Sacket et al. (1965)	$\delta^{13}\text{C}_{\text{POC}}$	not included	added	not included in Goericke (1994)
Wada et al. (1987)	$\delta^{13}\text{C}_{\text{POC}}$	not included	added	not included in Goericke (1994)
Fischer (1989)	$\delta^{13}\text{C}_{\text{POC}}$	Goericke (1994)	source value	rounded in Goericke (1994)
Fontugne and Duplessy (1978)	$\delta^{13}\text{C}_{\text{POC}}$	Goericke (1994)	source value	rounded in Goericke (1994)
Fontugne and Duplessy (1981)	$\delta^{13}\text{C}_{\text{POC}}$	Goericke (1994)	source value	rounded in Goericke (1994)
Fischer (1989)	temperature	Goericke (1994)	source value	rounded in Goericke (1994)
Fontugne and Duplessy (1981)	temperature	Goericke (1994)	source value	rounded in Goericke (1994)
Francois et al. (1993)	temperature	Goericke (1994)	source value	rounded in Goericke (1994)
Sacket et al. (1965)	temperature	Goericke (1994)	source value	rounded in Goericke (1994)
Fischer (1989)	$\delta^{13}\text{C}_{\text{POC}}$	Goericke (1994)	deleted	not found in source
Fontugne and Duplessy (1978)	temperature	Goericke (1994)	deleted	not found in source

* The original source was not available, but we highly suspected an error in the coordinates that interchanged east and west.

Table 2. Changes made in other data: this table's structure is equivalent that of Table 1. It refers to all changes made in general and any data other than the Goericke (1994) data.

Data set	Changed	From	To	Reason
Any	depth	“surface”	1	comparability
Any	depth	depth range	average ¹	comparability
Trull and Armand (2013a)	$\delta^{13}\text{C}_{\text{POC}}$	three available	“blank correction”	mentioned in Trull and Armand (2001)
Trull and Armand (2013b)	$\delta^{13}\text{C}_{\text{POC}}$	three available	“blank correction”	mentioned in Trull and Armand (2001)
Any using sediment traps	month, year	range	explicit value ²	comparability
Chang et al. (2013)	month, year	range	explicit number	just one date for trap sampling given
Lorrain	project/cruise		“campaign”	provided by Anne Lorrain
Tuerena	project/cruise		“source”	provided by Robyn Tuerena
Tuerena	long	[0°, 360°]	[-180°, 180°] ³	comparability
Lorrain	long	[0°, 360°]	[-180°, 180°] ³	comparability
MacKenzie et al. (2019)	depth	original	deleted	suspected typo
Voss and von Bodungen (2003)	trap duration	original	deleted	suspected typo
De Jonge et al. (2015a)	method	multiple investigations (MULT)	in situ pump	found in De Jonge et al. (2015b)

¹ By arithmetic mean. ² Only for sample durations entirely within an explicit month and year, otherwise information on time frames has been discarded. ³ We applied Eq. (2).

Table 3. Available data and meta-information: the columns of the raw data set correspond to the provided data and meta-information. Their names are given in the first column of this table. The second holds a short description of their content and the third their ranges of values. In the final column we give how well this data kind is covered relative to the size of the full data set.

Column	Content	Range of values	Coverage ¹
Reference	citation ²	description	full ³
No.	running index	{1, ..., 4732}	full
Lat	latitude in decimal degrees ⁴	[−90°, 90°]	4604/4732
Long	longitude in decimal degrees ⁴	[−180°, 180°]	4604/4732
d13C	$\delta^{13}\text{C}_{\text{POC}}^4$	[−55.15, −4.5]	full
d13Canomaly	$\delta^{13}\text{C}_{\text{POC}} - \text{mean}^5_{\delta^{13}\text{C}_{\text{POC}}}$	[−31.19, 19.46]	full
Temp	temperature in degrees Celsius ⁴	[−1.8, 31.12]	1622/4732
Month	month as number	{1, ..., 12}	4114/4732
Year	year CE	{1964, ..., 2015}	4483/4732
Depth	depth in meters	[0, 4850]	3917/4732
Method	measurement method of $\delta^{13}\text{C}_{\text{POC}}$	description	3164/4732
Origin	associated project or cruise	description	3921/4732
Note	special circumstances	description	140/4732
Trap duration	duration of trap activity in days	[1, 133]	533/587 ⁶

¹ Ratio of available entries relative to the full number of data points. ² Wherever possible, this includes: author(s), year, title, journal name, full, number, issue, pages and DOI. ³ Primary source was not available in every case as a reference. A note, where the data were taken is included in this case. ⁴ With as many decimal places as available.

⁵ Rounded to two decimal places. ⁶ Here, abundance is given relative to the full number of sediment trap samples.

as this, while no surface microlayer measurements were included. The month and year were used to describe the sample date; specific days are neglected.

Anomalies of $\delta^{13}\text{C}_{\text{POC}}$ are given in the δ ratio described in Eq. (1). A sample method was added, wherever available. Any special sampling circumstances were given in the “Note” column. Activity duration of sediment traps was denoted in the last column.

The “Origin” columns listed the associated project or cruise or author note. Some samples were given with multiple project connections; all of them were given in this column.

3.2 Interpolated data sets

The interpolated $\delta^{13}\text{C}_{\text{POC}}$ data are available as NetCDF files on two global grids with different resolutions. NetCDF files are machine-independent and support the creation, accessing and sharing of array-oriented scientific data. On the UVic grid, we provide seven different files, each of them independent of time and averaged over the available spatial information. Six of them contain an individual decade each (from the 1960s through the 2010s). The seventh file comprises a combined set of all interpolated $\delta^{13}\text{C}_{\text{POC}}$ data. On the WOA grid, we provide 13 files including all $\delta^{13}\text{C}_{\text{POC}}$ measurements with complete spatial–temporal information, averaged across time and space.

One major aim of this work is to support reliable validation and calibration of $\delta^{13}\text{C}_{\text{POC}}$ -simulating models. Hence, we chose the grid of the UVic model version 2.9, as used,

e.g., in Schmittner and Somes (2016). Horizontally, it consists of 100×100 cells with a resolution of $1.8^\circ \times 3.6^\circ$, arranged from 0 to 360° in longitude (LONG) and $−90$ to 90° in latitude (LAT). Vertically, it is split up into 19 vertical layers (DEPTH), decreasing in resolution with depth. The two uppermost layers reach down to depths of 50 and 130 m, respectively, and they are supposed to comprise the upper ocean’s euphotic zone.

The WOA grid is based on the $1^\circ \times 1^\circ$ grid of the World Ocean Atlas (Garcia et al., 2018). It has a horizontal resolution of 360 arranged from $−180$ to 180° in the longitude (LONG) and 180 arranged from $−90$ to 90° in the latitude (LAT) direction. Vertically, it is split up into 102 layers (DEPTH). The time axis (TIME) increases in increments for each year from 1964 to 2015 by 1 and has a size of 52. This interpolation includes only $\delta^{13}\text{C}_{\text{POC}}$ data with full spatio-temporal metadata coverage; i.e., additionally to latitude, longitude and depth, we also required and included year and month information.

Ferret scripts were used for the interpolations. These averaged the irregularly measured data points within the ocean grid to one single data point representing each covered grid cell. The interpolation function SCAT2GRIDGAUSS by NOAA’s Pacific Marine Environmental Laboratory (2020) performed the spatial averaging under PyFerret v7.5. Calculations in this function are based on a work by Kessler and McCreary (1992) and can be summarized as follows: let $(x_1, y_1), \dots, (x_n, y_n) \subseteq \mathbb{R}^2$ be an equidistant grid and $(\tilde{x}_1, \tilde{y}_1), \dots, (\tilde{x}_m, \tilde{y}_m) \subseteq \mathbb{R}^2$ be irregular measurement locations of a real tracer D_j , $j \in \{1, \dots, m\}$. Then the value $D_i \in$

\mathbb{R} at grid point $(x_i, y_i), i \in \{1, \dots, n\}$ becomes interpolated as

$$D_i := \frac{\sum_{j=1}^m D_j W_{i,j}}{\sum_{j=1}^m W_{i,j}}, \quad (4)$$

where

$$W_{i,j} := \begin{cases} 0; & \tau_{i,j} < e^{-CX} \\ 0; & \tau_{i,j} < e^{-CY} \\ \tau_{i,j} & \text{otherwise,} \end{cases} \quad (5)$$

where $\tau_{i,j} := \exp\left(-\left(\frac{(x_j-x_i)^2}{X^2} + \frac{(y_j-y_i)^2}{Y^2}\right)\right)$ is the Gaussian weight function, $X, Y \in \mathbb{R}$ comprises scaling arguments and $C \in \mathbb{R}$ the cutoff parameter. We set $X = 1.8$, $Y = 0.9$ and $C = 1$ in our script.

Since the interpolation into the WOA grid excluded all data without full spatio-temporal metadata coverage, we focus the following descriptions of interpolated data on the UVic grid interpolations. These also include data without month information in the six decadal files and even completely without temporal information in the seventh time-independent file.

4 Main data set characteristics

The final data set includes 4732 individual $\delta^{13}\text{C}_{\text{POC}}$ measurements of seawater samples. We show the distribution of $\delta^{13}\text{C}_{\text{POC}}$ values by Gaussian kernel density estimation (KDE) in Fig. 1. KDEs are non-parametric density estimations (Silverman, 1986) for the approximation of probability density functions, which are theoretically similar to histograms but with continuous curves not dependent on rigid intervals. We applied a Python implementation from the SciPy stats package (Virtanen et al., 2020) to create the results presented here. Likewise, we derived conditional probability densities of $\delta^{13}\text{C}_{\text{POC}}$ values, given the different measurement method applied (Fig. 3).

4.1 Range and outlier values

The data distribution is presented by its KDE in Fig. 1. The interval of $\delta^{13}\text{C}_{\text{POC}}$ values ranges over $[-55.15, -4.5]$ with a mostly smooth distribution. Most of our data exhibit values around $\delta^{13}\text{C}_{\text{POC}} \approx -24\text{\textperthousand}$, which becomes clearly identifiable as a single maximum in the KDE. Two smaller modes are visible at around $\delta^{13}\text{C}_{\text{POC}} \approx -27.5\text{\textperthousand}$ and $\delta^{13}\text{C}_{\text{POC}} \approx -22\text{\textperthousand}$ (see also Table A1 in the Appendix). A steep decline to zero is visible outside the two outer modes. The steep decline in the KDE stops at around $\delta^{13}\text{C}_{\text{POC}} = -37\text{\textperthousand}$ and $\delta^{13}\text{C}_{\text{POC}} \approx -14\text{\textperthousand}$. Between $\delta^{13}\text{C}_{\text{POC}} \approx -37\text{\textperthousand}$ and $\delta^{13}\text{C}_{\text{POC}} \approx -55.15\text{\textperthousand}$ as well as between $\delta^{13}\text{C}_{\text{POC}} \approx -14\text{\textperthousand}$ and $\delta^{13}\text{C}_{\text{POC}} \approx -4.5\text{\textperthousand}$ the KDE closely aligns to the x axis, which indicates very few data points lie in this range.

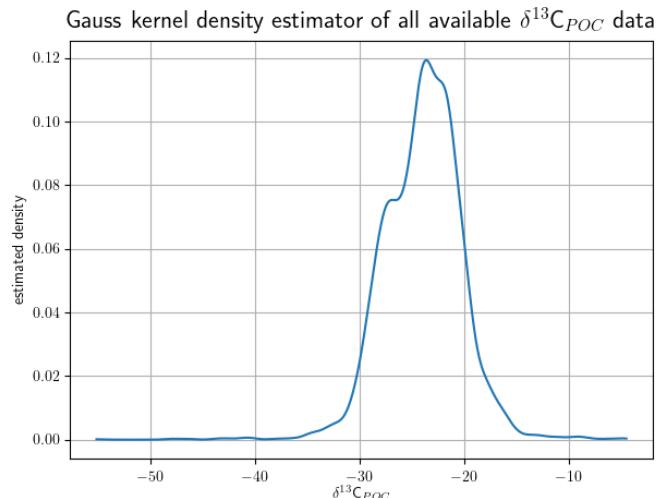


Figure 1. The density function of all individual $\delta^{13}\text{C}_{\text{POC}}$ measurements approximated by Gaussian kernel density estimation: values of the estimated density are drawn on the y axis; the $\delta^{13}\text{C}_{\text{POC}}$ values run on the x axis. The higher the value of the estimated density is, the more $\delta^{13}\text{C}_{\text{POC}}$ points have been measured around this value.

Below $\delta^{13}\text{C}_{\text{POC}} = -37\text{\textperthousand}$ we find 17 data points ranging down to $\delta^{13}\text{C}_{\text{POC}} = -55.15\text{\textperthousand}$. Down to $\delta^{13}\text{C}_{\text{POC}} = -48\text{\textperthousand}$ these were all taken from Lein and Ivanov (2009) and Lein et al. (2006), measured in September or October 2003, around the location $10^\circ\text{N}, 104^\circ\text{W}$ and below 2500 m depth in the vicinity a hydrothermal field close to the Pacific coast of middle America. The lowest outlier at $\delta^{13}\text{C}_{\text{POC}} = -55.15\text{\textperthousand}$ was taken from Altabet and Francois (2003a) from November 1996 and at $62.52^\circ\text{S}, 169.99^\circ\text{E}$ at the ocean surface south of New Zealand.

Above $\delta^{13}\text{C}_{\text{POC}} = -10\text{\textperthousand}$ we find 15 data points ranging up to $\delta^{13}\text{C}_{\text{POC}} = -4.5\text{\textperthousand}$. Three of them were taken from Lein et al. (2007) and measured at 800 m depth at a hydrothermal vent located $30.125^\circ\text{N}, 42.117^\circ\text{W}$ in the middle north Atlantic. Ten were taken from Calvert and Soon (2013b, c, a). All of these were measured between 636 and 901 m depth around $49^\circ\text{N}, 130^\circ\text{W}$ close to the American coast of the Pacific, and all of them were measured in February or May, except one in August. The final two were part of the Lorrain data set. Both were measured at the ocean surface in the South Pacific, in July at $5.3^\circ\text{S}, 164.9^\circ\text{E}$ and December at $20.9^\circ\text{S}, 159.6^\circ\text{E}$.

Since more than 98 % of the data (4668 of the 4732 data points) have values that lie between $\delta^{13}\text{C}_{\text{POC}} = -35\text{\textperthousand}$ and $\delta^{13}\text{C}_{\text{POC}} = -15\text{\textperthousand}$, we will focus on this range in our following analyses.

We tested the robustness of our KDE approach in a subsampling experiment. We considered 500 random subsets of 20 % of the original data over the range with the highest data density $[-35, -15]$ and visualize their KDEs in Fig. 2. They show peaks at $\delta^{13}\text{C}_{\text{POC}} \approx -23\text{\textperthousand}$, fitting the maximum and the second smaller mode to the right of it, and

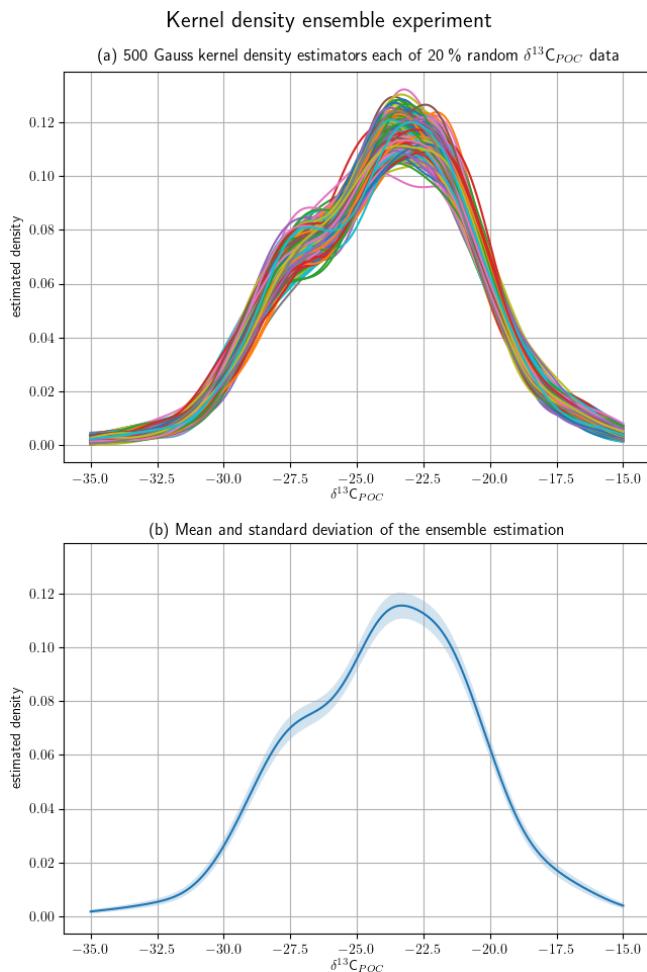


Figure 2. A random sample of 20 % of the $\delta^{13}\text{C}_{\text{POC}}$ data was taken from the full data set for 500 times to generate an ensemble of subsets. Their densities were approximated with a Gaussian kernel density estimator. Panel (a) shows all 500 estimated densities by individual lines. Panel (b) shows the mean and the variance of the full ensemble of densities by a graph and the shaded area around it, respectively.

at $\delta^{13}\text{C}_{\text{POC}} \approx -27.5\text{\textperthousand}$. Outside $[-27, -22]$ the KDEs are closely aligned. The mean of and standard variation in the KDE ensemble also show the highest variability around the two modes at $\delta^{13}\text{C}_{\text{POC}} = -23\text{\textperthousand}$ and $\delta^{13}\text{C}_{\text{POC}} = -27.5\text{\textperthousand}$.

4.2 Sampling methods

Various sampling methods were involved in obtaining the $\delta^{13}\text{C}_{\text{POC}}$ data. Around 67 % of the data had associated sampling-method information, which included 18 different sampling methods. In principle, all 18 methods could be grouped into five main observational types: bottles, intake, nets, traps and diverse. ‘‘Bottle’’ data include samples taken from Niskin bottles and samples collected via Sea-Bird submersible pumps. By ‘‘intake’’ we refer to all versions of

pumps and underway cruise track measurements, as well as multiple-unit large-volume filtration systems (MULVFSs). ‘‘Net’’ data represent all occurring versions of plankton nets, and ‘‘traps’’ refers to all represented sediment traps and moorings. Finally, the deep-sea manned submersible (MIR2) is not classified into any of these groups and was assigned to a cluster that we refer to as ‘‘diverse’’.

All sample devices provided data over all sample depths. Deeper samples were mainly taken from traps and pump systems and the upper samples from bottle and net data. Most data sampled deeper than 2600 m were collected by sediment traps. At 3800 m there were several trap contributions by Calvert (e.g., Calvert, 2002), mostly from the late 1980s. Data sampled by a deep-sea manned submersible were given at locations down to 2520 m (Lein and Ivanov, 2009).

For resolving differences between sampling methods we chose data from the Atlantic Ocean which comprise all four major methods (with data embracing a region between 45° S and 80° N and 70° W and 20° E). In addition, data were distinguished by tropical, temperate and polar subregions. By crudely sorting the data according to their sampling locations, we gain some insight into methodological variability within a subregion and may relate this to variations between the three subregions (Fig. 3). Overall, we do not find any severe bias with respect to any particular method. Bottle data seem to cover most of the lower $\delta^{13}\text{C}_{\text{POC}}$ values that typically range between $-28\text{\textperthousand}$ and $-21\text{\textperthousand}$, which could be due to samples collected at greater depths. Intake and net measurements are rather restricted to the upper ocean layers, and these methods often yield $\delta^{13}\text{C}_{\text{POC}}$ values larger than $-25\text{\textperthousand}$, with some polar net measurements being a notable exception (Fig. 3d). For the tropical Atlantic (30° S– 30° N) the net and intake measurements vary around $-21\text{\textperthousand}$, with 95 % confidence limits between $-24\text{\textperthousand}$ and $-18\text{\textperthousand}$ (see Table A2 in the Appendix). According to our comparison, we could not identify any method that yields much greater variance of $\delta^{13}\text{C}_{\text{POC}}$ values than others. The spatio-temporal variations of the $\delta^{13}\text{C}_{\text{POC}}$ compare well among different methods, but we advise caution when comparing bottle measurements with data of other methods because of potential differences in the depth range covered.

In the full Atlantic Ocean, densities of intake and net data are most representative of the maximum full $\delta^{13}\text{C}_{\text{POC}}$ sample. From the intake data shown here, $\sim 80\text{ \%}$ were sampled within 30° S and 30° N. When restricting data to this area, net data better resemble the full data. Of all net sample data, $\sim 80\text{ \%}$ were collected between 30 and 60° N, where they fit the overall $\delta^{13}\text{C}_{\text{POC}}$ density best, followed by trap data. Trap and bottle data deliver the lowest $\delta^{13}\text{C}_{\text{POC}}$ measurements in the Atlantic Ocean. Of both data kinds, $\sim 74\text{ \%}$ to 85 \% were sampled north of 60° N. A restriction to this area shows trap and bottle samples being closely aligned to the full data in this region.

The variance of the intake and trap data is $\sim 3\text{\textperthousand}$ and lower than the variance of all $\delta^{13}\text{C}_{\text{POC}}$ together, which is

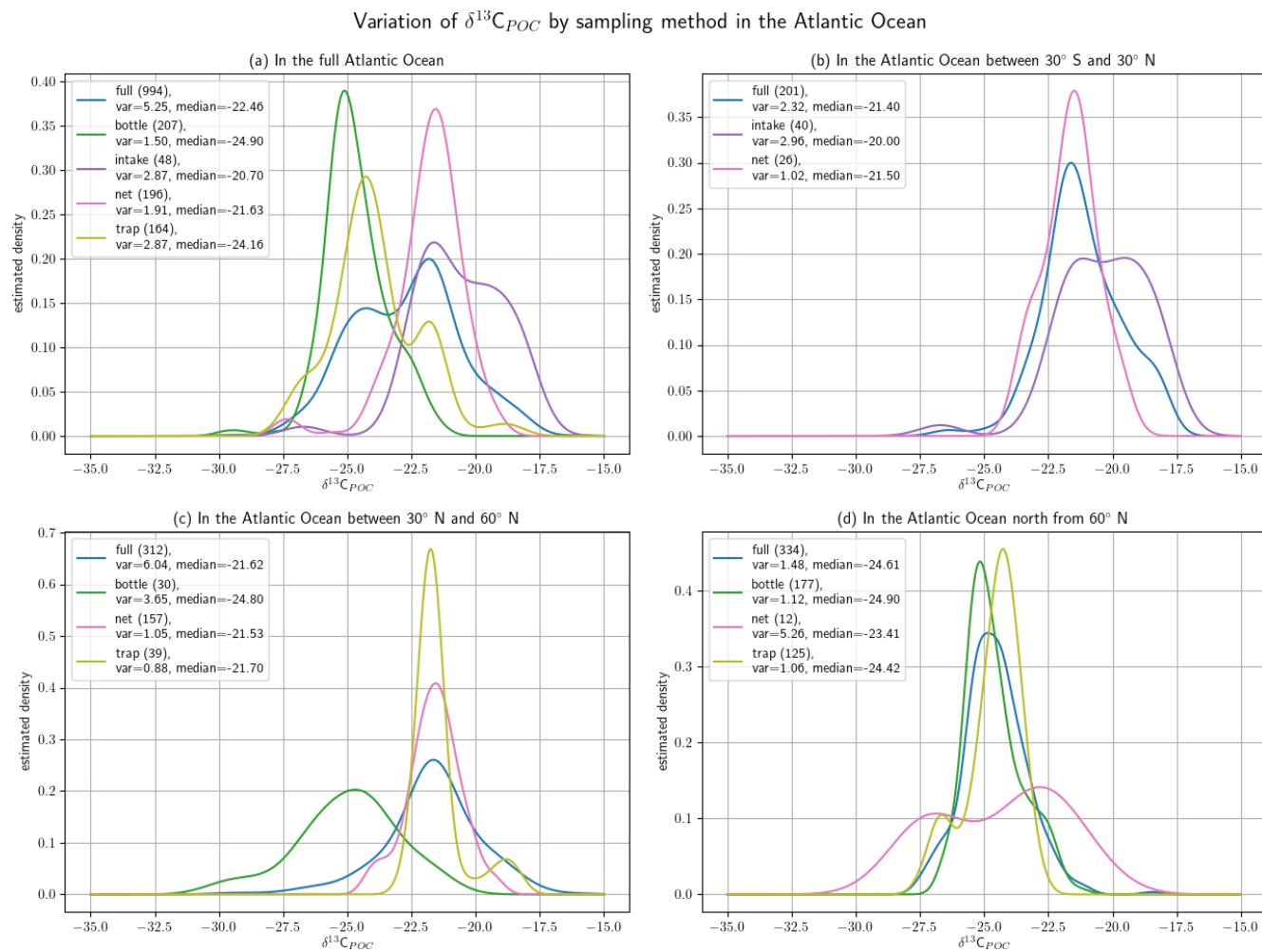


Figure 3. Separation of $\delta^{13}\text{C}_{\text{POC}}$ in the Atlantic Ocean data by four main sample methods: bottle, intake net and trap data. Panel (a) shows the full Atlantic Ocean, panel (b) the equatorial core of the Atlantic Ocean, panel (c) the Atlantic between 30° S and 30° N, and panel (d) its most northern area. In each plot, the density of the $\delta^{13}\text{C}_{\text{POC}}$ sample groups with enough data was approximated by Gaussian KDEs and drawn with an individual color. An additional graph shows the comparison to the full- $\delta^{13}\text{C}_{\text{POC}}$ -data density in the respective area. The numbers of used data points are indicated in each KDE label.

~5‰, the highest value observed here. Both bottle and net data show a variance of less than 2‰. Furthermore, trap, net and full $\delta^{13}\text{C}_{\text{POC}}$ show a pronounced second mode in their densities, while bottle and net data show a clear individual maximum. Median values of net and intake data are ~1‰ to ~2‰ higher, respectively, than the one of the full data. This has a median of $\delta^{13}\text{C}_{\text{POC}} = 22.46\text{\textperthousand}$. Both bottle and trap data show a ~2‰ lower median. Analytical errors and uncertainties are typically 0.2‰ or lower (Young et al., 2013) and thus are not likely to significantly contribute to the much larger variance in the observations

5 Spatial distribution

We show the spatial distribution of $\delta^{13}\text{C}_{\text{POC}}$ measurements across the global ocean surface and depths. Most $\delta^{13}\text{C}_{\text{POC}}$

data have been measured in the uppermost few ocean meters, and the best surface coverage is available for the Atlantic Ocean. Changes in $\delta^{13}\text{C}_{\text{POC}}$ on the ocean surface were evaluated based on the UVic grid.

5.1 Vertical distribution of the data set

Depth values are available for more than 80 % of the sample data with most of them located in the upper ocean. The distribution of depth measurements is shown in Fig. 4. An approximation of the depth measurements by Gaussian KDE is visualized in Fig. 5 along with the $\delta^{13}\text{C}_{\text{POC}}$ value distribution over them in the main ocean basins. The KDE resolves the best data coverage for the uppermost ~500 m of the oceans and a second far smaller maximum at ~3800 m. The depth ranges presented in Fig. 4 correspond to the depth intervals of

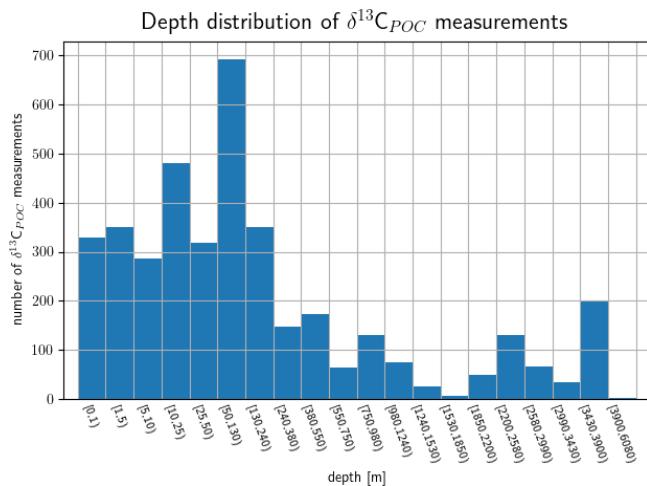


Figure 4. Vertical data coverage in depth layers based on the UVic grid: the uppermost 50 m is divided into subranges; below they are according to the UVic grid. The number of $\delta^{13}\text{C}_{\text{POC}}$ data points available is plotted against its respective depth range.

the UVic grid; only the two uppermost layers are presented in more detail, and the last four are combined. Within the first 130 m we observe the highest data density and find nearly 2500 measurements of $\delta^{13}\text{C}_{\text{POC}}$, where nearly 1000 of them were measured within [0 m, 10 m]. A total of 200 $\delta^{13}\text{C}_{\text{POC}}$ values were available in the depth interval [3430 m, 3900 m]. The two deepest values were taken from Fischer (1989) and Altabet and Francois (2003b) and sampled at 4500 and 4850 m depth, respectively.

Values of $\delta^{13}\text{C}_{\text{POC}}$ are, apart from in the North Pacific, closely aligned within the individual ocean basins. The Atlantic, South Pacific and Indian Ocean show values mostly of $-28\text{\textperthousand}$ to $-19\text{\textperthousand}$. The $\delta^{13}\text{C}_{\text{POC}}$ values in the Arctic reach down to approximately $-30\text{\textperthousand}$ and those in the Southern Ocean even to approximately $-35\text{\textperthousand}$. The North Pacific shows a wide spread of $\delta^{13}\text{C}_{\text{POC}}$ values, especially between 50 and 100 m depth and at 2500 m depth. There they reach either down to less than $-40\text{\textperthousand}$ or up to more than approximately $-10\text{\textperthousand}$ at a depth of 2500 m.

Measurements in the North Atlantic, North Pacific and Indian Ocean reach down to more than 3500 m. Measurements down to nearly 5000 m were sampled in the Southern Ocean. The South Pacific was sampled down to a depth of 2500 m and the Arctic Ocean and South Atlantic only in the uppermost few hundred meters.

5.2 Horizontal distribution of the data set

All global oceans are covered with $\delta^{13}\text{C}_{\text{POC}}$ data. In Fig. 6 the horizontal distribution of available data is depicted for both grids. For the UVic grid we show data from the file including all data independent of time; the WOA grid is averaged over all times. In both cases, we averaged data over all

depths and also added data without depth information to best visualize the horizontal coverage. A similar plot, although with a different purpose, is given later in this work in Fig. 10 showing only surface data locations.

Many cruises are visible as lines formed by connected grid cells in Fig. 6, especially in the Atlantic and Indian Ocean and less so in the Southern Ocean. Also, smaller sample spots occur, mainly located in the Pacific, Arctic and Southern Ocean. The Atlantic Ocean provides the best data coverage. Then the Southern and Indian oceans contain the next best coverage with the North Pacific having the sparsest.

The highest $\delta^{13}\text{C}_{\text{POC}}$ values are evident in low-latitude regions. In the Atlantic Ocean the highest values were measured between $0\text{--}30^\circ\text{N}$ and $30\text{--}60^\circ\text{W}$ as well as close to the western coast of France, reaching up to at least $-17\text{\textperthousand}$. The Indian Ocean shows generally high values of approximately $-20\text{\textperthousand}$. In the Pacific Ocean the highest values are close to the Peruvian coast and Papua New Guinea. We also find high values in the Bering Strait and on the northern edge of the Southern Ocean at around 65°E .

The lowest $\delta^{13}\text{C}_{\text{POC}}$ values are mostly found in the Southern Ocean. Nearly all measured grid cells here belong to $\delta^{13}\text{C}_{\text{POC}}$ values lower than around $-28\text{\textperthousand}$. The Arctic Ocean shows low values as well, for instance in the Kara Sea. The lowest values in the Pacific Ocean occur in the Southern Ocean at high latitudes.

5.3 Meridional trend of $\delta^{13}\text{C}_{\text{POC}}$ values

We show the north–south trend of $\delta^{13}\text{C}_{\text{POC}}$ over the Atlantic Ocean based on the time-independent UVic grid and restricted to the uppermost 130 m, which resembles the euphotic zone in the UVic model. We chose this section due to it having the best data coverage. A biome mask according to Fay and McKinley (2014) was applied to the gridded data, thereby defining latitudinal zones in the entire Atlantic Ocean. Distributions of $\delta^{13}\text{C}_{\text{POC}}$ within the biomes are shown in Fig. 7 (see also Table A3 in the Appendix).

The biomes derived by Fay and McKinley (2014) are areas with consistent biological and ecological properties. The chosen biomes cover the Atlantic Ocean and extend to the Arctic Sea and parts of the Southern Ocean. The biomes are numbered 9 to 17, excluding 14. The biomes 15 to 17 represent parts of the Southern Ocean and were restricted to 70°W and 20°E . Their locations are shown in Fig. 7.

Observations by the biomes are consistent with the ones from Fig. 6. The two biomes showing the lowest $\delta^{13}\text{C}_{\text{POC}}$ values from $-28\text{\textperthousand}$ to $-29\text{\textperthousand}$ are those located farthest south. The biome located farthest north contains the next-lowest values of about $-24\text{\textperthousand}$. The biomes with more positive $\delta^{13}\text{C}_{\text{POC}}$ values are in the lower latitudes and show similarly higher values from $-23\text{\textperthousand}$ to $-21\text{\textperthousand}$.

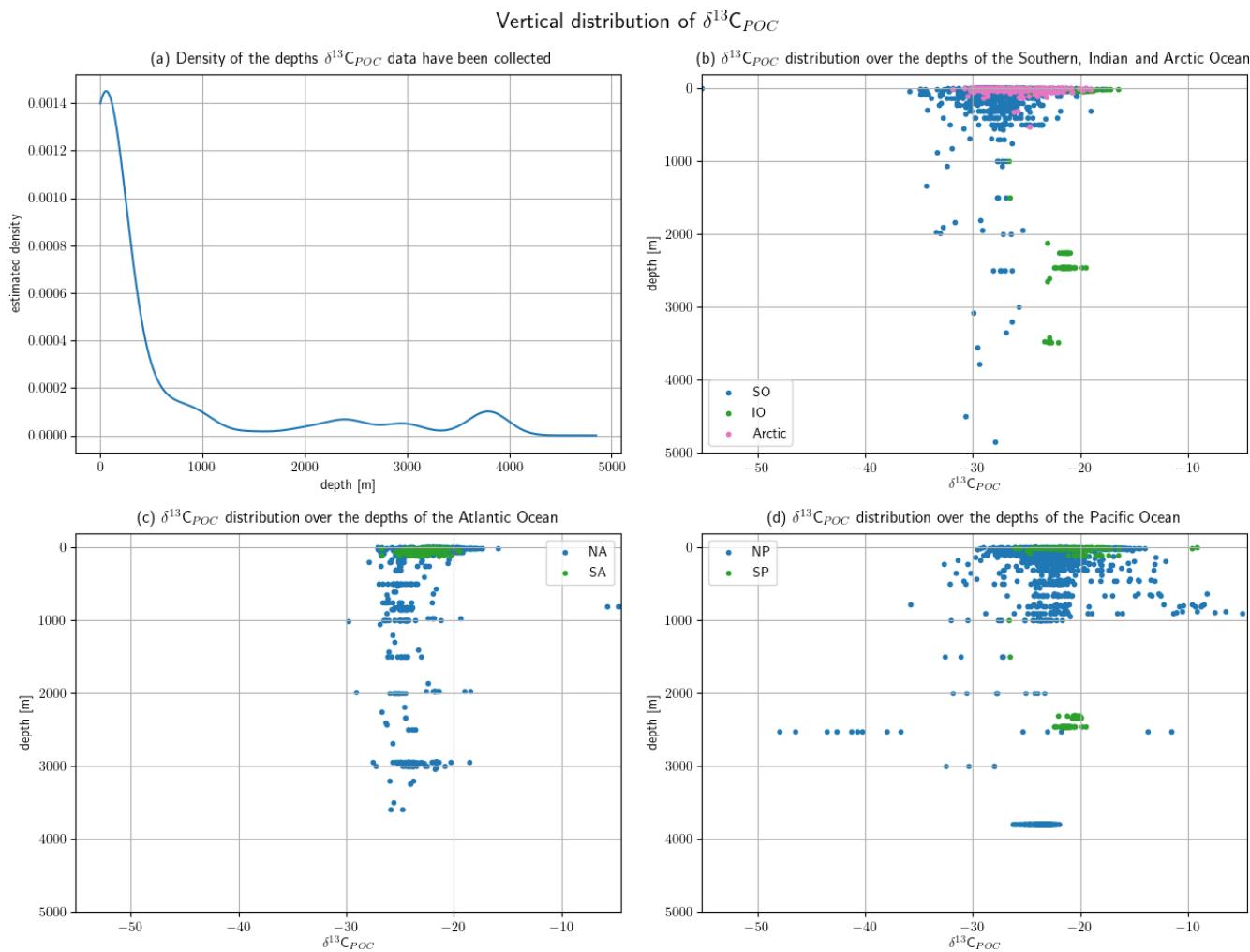


Figure 5. The vertical distribution of available $\delta^{13}\text{C}_{\text{POC}}$ samples is shown (**a**) as the approximated density of the measurement depths and (**b–d**) as measured $\delta^{13}\text{C}_{\text{POC}}$ values relative to their respective measurement depth. Panel (**a**) provides the estimated density of the depth values on the y axis and the depth in meters on the x axis. The estimation was realized by a Gaussian KDE. Panel (**b**) resolves the measurements of the Southern, Indian and Arctic Ocean, (**c**) the North Atlantic and South Atlantic, and (**d**) the North Pacific and South Pacific. The last three panels show the depth in meters on the y axis and the measured $\delta^{13}\text{C}_{\text{POC}}$ value on the x axis. Different colors are used to mark different ocean basins.

6 Temporal distribution of the data set

The full $\delta^{13}\text{C}_{\text{POC}}$ data cover a time period of around 50 years over 1964–2015 and all 12 calendar months. The number of samples measured during individual decades varies considerably with most measurements in the 1990s. Coverage within the months is quite comparable; only winter months in both hemispheres exhibit fewer data.

The distribution of $\delta^{13}\text{C}_{\text{POC}}$ samples over the years is resolved in Table 4 and is visually approximated by Gaussian KDE in Fig. 8. The 1990s shows the best data coverage. More than half of the data points are associated with a year in this decade, which is visible by a pronounced maximum in the estimated density. The sparsest data are found in the 1960s, when only 74 data points were sampled. All other decades

come with between around 300 and 600 $\delta^{13}\text{C}_{\text{POC}}$ data points. The latest data are mostly from Anne Lorrain, MacKenzie et al. (2019) and Kaiser et al. (2019). The oldest data were taken from the data sets by Robyn Tuerena, Degens et al. (1968), and Eadie and Jeffrey (1973).

6.1 Monthly variations

Monthly clustered data of the Northern Hemisphere and Southern Hemisphere show monthly variations, but more observations are required to demonstrate robust seasonality within different regions. Since more than 50 % of the available $\delta^{13}\text{C}_{\text{POC}}$ data originate in the 1990s, we selected data from this decade to exclude changes that might be introduced by longer-term trends. Furthermore, we restricted our data

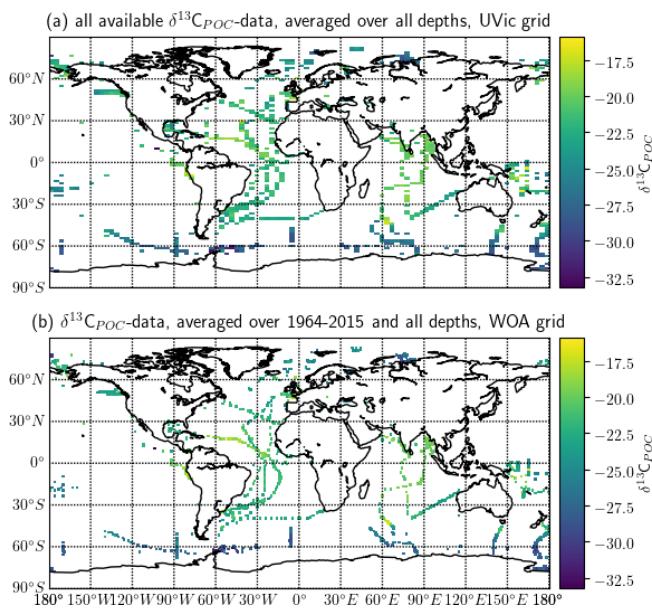


Figure 6. Global distribution of the $\delta^{13}\text{C}_{\text{POC}}$ data is visualized based on the (a) UVic grid and (b) WOA grid. The data used for (a) are independent of time and include all available measurements with latitude and longitude information. The data shown in (b) include only data with complete temporal metadata and are averaged over the years 1964–2015. Both kinds of data are averaged over all measurements including data with missing depth information. Each colored square refers to a grid cell with available $\delta^{13}\text{C}_{\text{POC}}$ measurements. The colors indicate the $\delta^{13}\text{C}_{\text{POC}}$ value in the respective grid cell.

Table 4. Data coverage within the available decades: the first column lists the available decades and the second column the number of sampled $\delta^{13}\text{C}_{\text{POC}}$ data points within this time frame.

Decade	$\delta^{13}\text{C}_{\text{POC}}$ values available
1960s	74
1970s	321
1980s	463
1990s	2403
2000s	614
2010s	589

to the uppermost 130 m, which resembles the euphotic zone in the UVic model. In Fig. 9 we displayed all months with enough data points by a KDE and indicate the same months by the same colors. We excluded July, November and December in the Northern Hemisphere from this KDE representation because these months provided three or fewer data points each, which resulted in a KDE that overgrew the others by magnitudes and made their visual comparison difficult. The KDEs are supported by comparison of the median values of the individual months in Table 5.

The monthly resolved variations in $\delta^{13}\text{C}_{\text{POC}}$ do not reveal any significant seasonal pattern (Fig. 9; see also Table A4 in the Appendix). In general we find the highest $\delta^{13}\text{C}_{\text{POC}}$ values in the Northern Hemisphere, with a median $\delta^{13}\text{C}_{\text{POC}}$ of $-20.4\text{\textperthousand}$ in April and a median $\delta^{13}\text{C}_{\text{POC}}$ of $-21.5\text{\textperthousand}$ in October, which are typical months with enhanced primary production (Northern Hemisphere spring and autumn blooms). Similarly high median $\delta^{13}\text{C}_{\text{POC}}$ values cannot be ascertained for any month with data of the Southern Hemisphere, where values of $\delta^{13}\text{C}_{\text{POC}}$ above $-20\text{\textperthousand}$ have rarely been observed at any time of the year. In fact, there is an overall tendency towards low $\delta^{13}\text{C}_{\text{POC}}$ values for the Southern Hemisphere, which becomes well expressed during the months April and September, with medians of $\delta^{13}\text{C}_{\text{POC}} = -28.1\text{\textperthousand}$ and $\delta^{13}\text{C}_{\text{POC}} = -28.5\text{\textperthousand}$, respectively. However, interpretations of this north–south trend should be treated with caution because the apparent tendency is likely conditioned by some imbalance in the number of high-latitude data points. Compared to the number of data points from the Southern Ocean, samples from the Arctic Ocean are considerably underrepresented (see also Fig. 10). Furthermore, the discrimination between data of the Northern Hemisphere and Southern Hemisphere is crude, and we encourage the use of our data collection for more advanced analyses of seasonal, monthly based changes in the $\delta^{13}\text{C}_{\text{POC}}$ signal.

6.2 Decadal variations

The decadal UVic grid NetCDF files are the basis for showing long-term changes in the $\delta^{13}\text{C}_{\text{POC}}$ data. An overview of where the data within the individual decades were sampled is given in Fig. 10. This shows that the sparsest coverage was obtained in the 1960s, located close to the central American continent. Most data in the Indian Ocean were sampled in the 1970s. A cruise across the southern part of the Atlantic Ocean up to 30° N and some samples close to Iceland were also measured in this decade. The 1980s is similarly sparse in spatial coverage to the 1960s. Measurements of the 1980s were taken at locations in the Southern Ocean, in the Arctic and in the Atlantic close to the Equator. The 1990s has the best coverage including most ocean basins. Most Southern Ocean data were sampled within the 1990s. The 2000s provides good coverage of the Arctic Ocean. Finally, the 2010s data were mostly sampled in the Southern Hemisphere in the open Pacific and Atlantic. A smaller number of Eurasian continental sea data were also part of the 2010s samples.

We show the changes in $\delta^{13}\text{C}_{\text{POC}}$ values over the available decades by density estimates in Fig. 11 (see also Table A5 in the Appendix) and by their median in Fig. 12. Figure 11 visualizes the sparse coverage of the Southern Ocean outside of the 1990s, which is why the area is not part of any further discussion here. The Southern Ocean is defined as the ocean area south of 45° S . All presented analyses were restricted to the euphotic zone, i.e., the uppermost 130 m resembling the two first layers of the UVic grid.

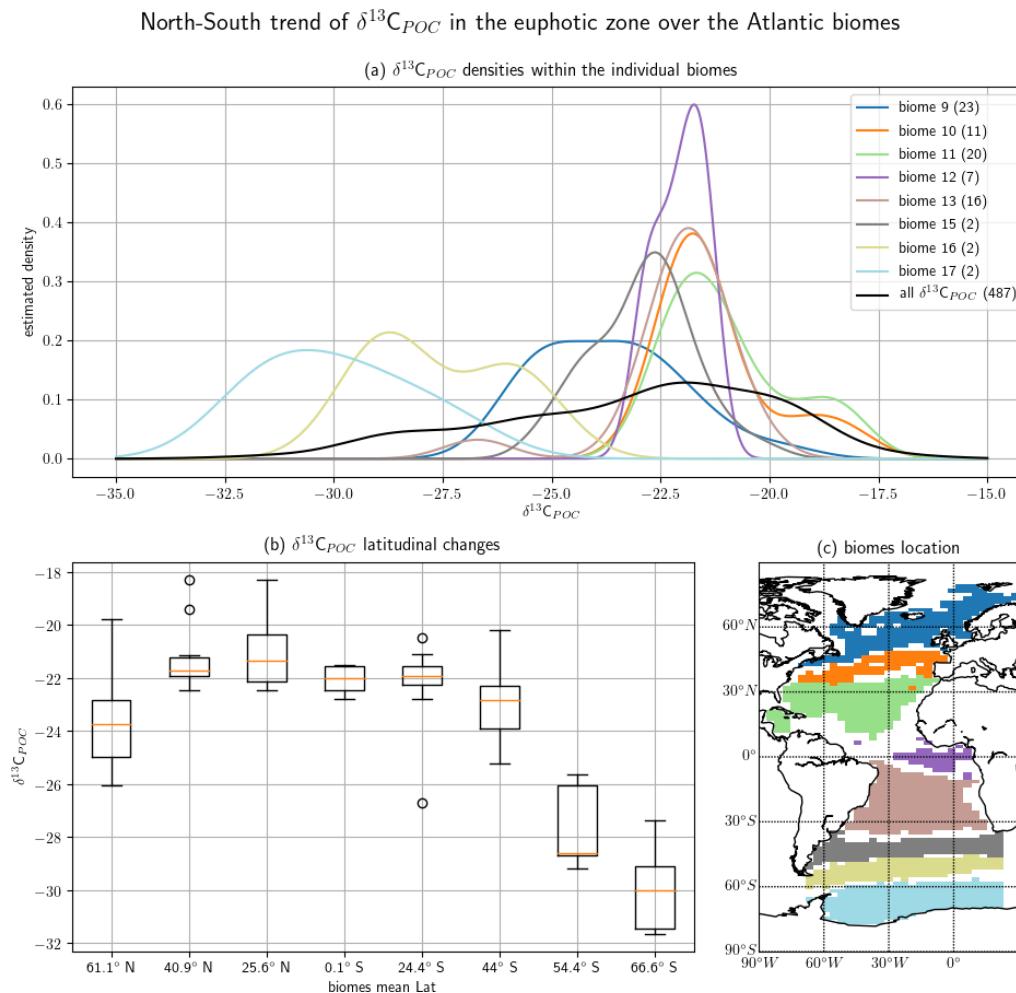


Figure 7. The north–south trend of sampled $\delta^{13}\text{C}_{\text{POC}}$ values is visualized by a cross section over the Atlantic Ocean. Biomes (Fay and McKinley, 2014) define the latitudinal bands of the interpolated data set. Panel (a) presents a Gaussian KDE for each biome approximating the density of the contained $\delta^{13}\text{C}_{\text{POC}}$ data. Different colors mark the individual biomes, and a black line shows the general global $\delta^{13}\text{C}_{\text{POC}}$ distribution. The number in parentheses in each KDE label counts the number of $\delta^{13}\text{C}_{\text{POC}}$ measurements used for the respective graph. Panel (b) shows in a box plot the steep decline in $\delta^{13}\text{C}_{\text{POC}}$ values from the tropical biomes towards the higher latitudes. The x axis provides the mean latitudes of the biomes introduced in (a). The y axis measures the $\delta^{13}\text{C}_{\text{POC}}$ value. Panel (c) shows the biome locations. Each biome is drawn in the color of its corresponding density estimate in (a) above. The biome numbers increase from the north to the south.

Table 5. Monthly median change in $\delta^{13}\text{C}_{\text{POC}}$. Due to their having the best data coverage, the analyses were carried out within the 1990s and in the uppermost 130 m.

Hemisphere	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
North		-24.815	-24.12	-20.	-24.06	-24.7	-21.746	-23.67	-22.83	-21.4	-23.5455	-23.368
South	-26.45	-26.41	-23.34	-28.2				-28.65	-27.95	-27.9	-26.08	

A clear decrease in $\delta^{13}\text{C}_{\text{POC}}$ densities in Fig. 11 can be identified for the global ocean outside of the Southern Ocean. All decades but the 1980s show one clear maximum in their approximated densities. The 1980s shows a second expressed density maximum at lower values. The main maximum shifts from the 1960s at $\delta^{13}\text{C}_{\text{POC}} \approx -19.9\text{\textperthousand}$ to the 2010s at $\delta^{13}\text{C}_{\text{POC}} \approx -23\text{\textperthousand}$. This decrease is also clearly

visible in the comparison of the decadal medians (Fig. 12). The Southern Ocean provides far worse data coverage. Only the 1980s and 1990s include enough data to construct a comparable KDE. Due to this very low data availability, all of these results must be taken with the highest caution.

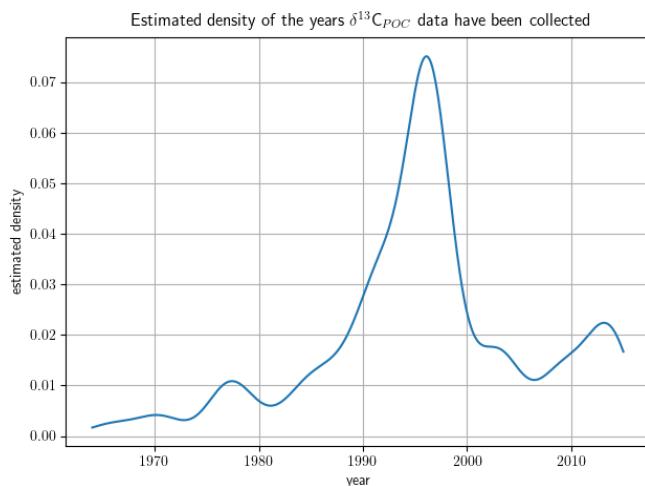


Figure 8. The distribution of $\delta^{13}\text{C}_{\text{POC}}$ data samples over the years approximated by Gaussian KDE. The density is drawn on the y axis; the sample year is on the x axis. A higher altitude of the graph indicates years with more available data.

7 Data availability

The described $\delta^{13}\text{C}_{\text{POC}}$ data are available at <https://doi.org/10.1594/PANGAEA.929931> (Verwega et al., 2021).

8 Conclusions

The aim of this work was to construct the largest publicly accessible $\delta^{13}\text{C}_{\text{POC}}$ data set. The starting point of our collection and analyses was the readily available data collection of Goericke (1994), which comprised 467 data points. Our primary objective was to elaborate this set of data by adding useful meta-information from the original publications and by introducing additional $\delta^{13}\text{C}_{\text{POC}}$ measurements, as recorded in the world ocean database PANGAEA and made available by Robyn Tuerena and Anne Lorrain. This way we could expand the data collection substantially, from the original 467 to 4732 data points. This new $\delta^{13}\text{C}_{\text{POC}}$ data set provides the best coverage to date and will be a useful tool to help constrain many marine carbon cycling processes and pathways from ocean–atmosphere exchange to marine ecosystems, as well as to better understand observations and validate models. To ensure dynamic growth of our data collection, the corresponding author will provide annual updates of the data set. Furthermore, he may be contacted by any interested researcher who would like to add their data to this collection.

The data are provided in a csv structure and interpolated onto two different global grids in NetCDF format. The csv file contains the $\delta^{13}\text{C}_{\text{POC}}$ values, their anomalies to their mean and all available meta-information. The interpolations are provided on a coarse $1.8^\circ \times 3.6^\circ$ grid of a $\delta^{13}\text{C}_{\text{POC}}$ -simulating model and a finer $1^\circ \times 1^\circ$ grid by the World Ocean Atlas. We have provided a detailed description of our data

Monthly variation of $\delta^{13}\text{C}_{\text{POC}}$ in the euphotic zone in the 1990s

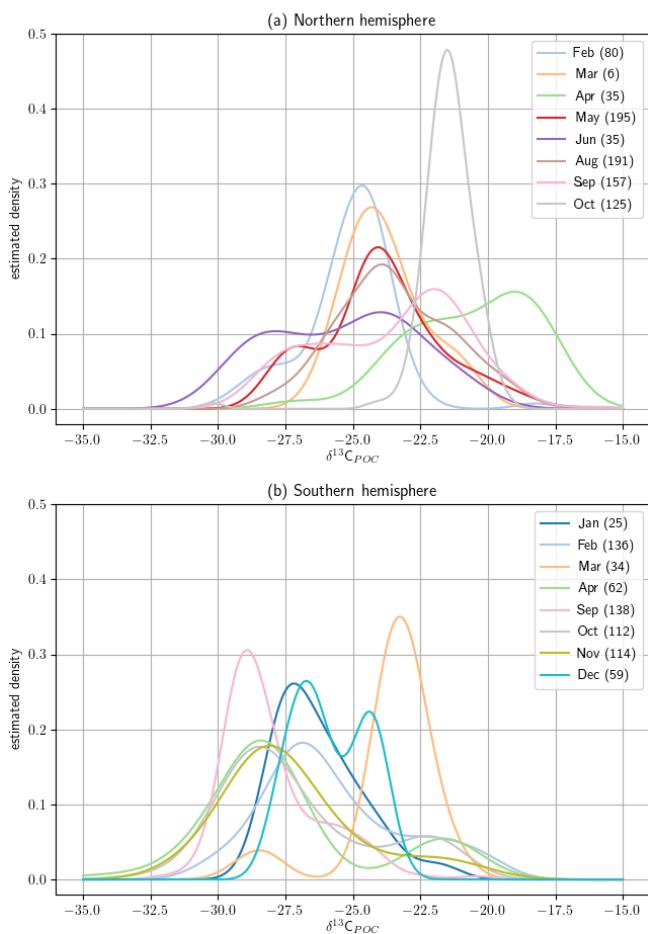


Figure 9. Monthly variations are split up by hemisphere with the Northern Hemisphere in **(a)** and Southern Hemisphere in **(b)**. Due to their having the best data coverage, the analyses are carried out within the 1990s and in the uppermost 130 m. The $\delta^{13}\text{C}_{\text{POC}}$ values are split up by sample month, and for every month with enough available data points (here more than three) a Gaussian KDE approximates their density. The number of used data points is given in each KDE label. For each hemisphere the densities are drawn all together; each month is indicated by an individual color.

collection procedure, all added meta-information and data coverage as well of the interpolation procedure carried out. We took the utmost care to make all data coherent, comparable and back-trackable and all adjustments transparent. Assumptions, changes and deletions of the used data sets have been described in detail.

We have described the general spatial and temporal trends of the sampled $\delta^{13}\text{C}_{\text{POC}}$ data of the raw data file. Distributions were always approximated by Gaussian kernel density estimators. The data range from 1964–2015 with by far the best coverage in the 1990s. Sample locations reach down to a depth of nearly 5000 m and best cover the uppermost 10 m, especially in the Atlantic and Indian Ocean. We were able to

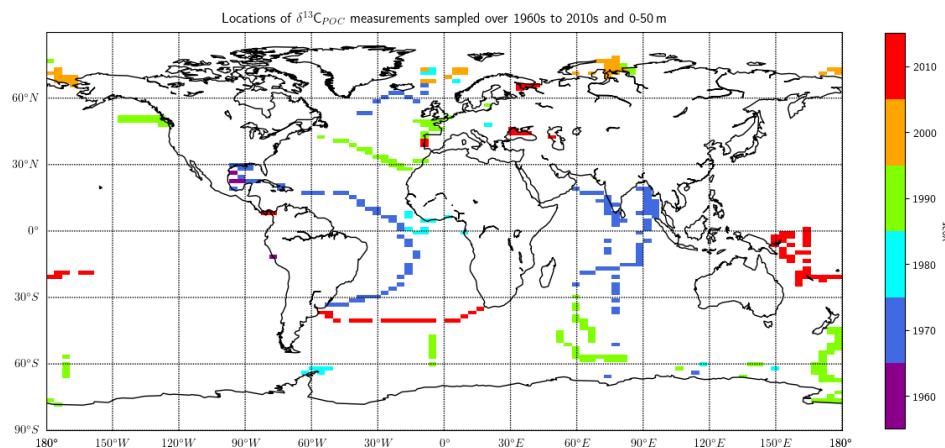


Figure 10. Grid locations of the $\delta^{13}\text{C}_{\text{POC}}$ data, colored by sampling decades. Only data of the uppermost layer are considered in this plot. The different colors indicate the different sample decades and were plotted increasing in time above each other.

Temporal shift of $\delta^{13}\text{C}_{\text{POC}}$ in the euphotic zone over last six decades

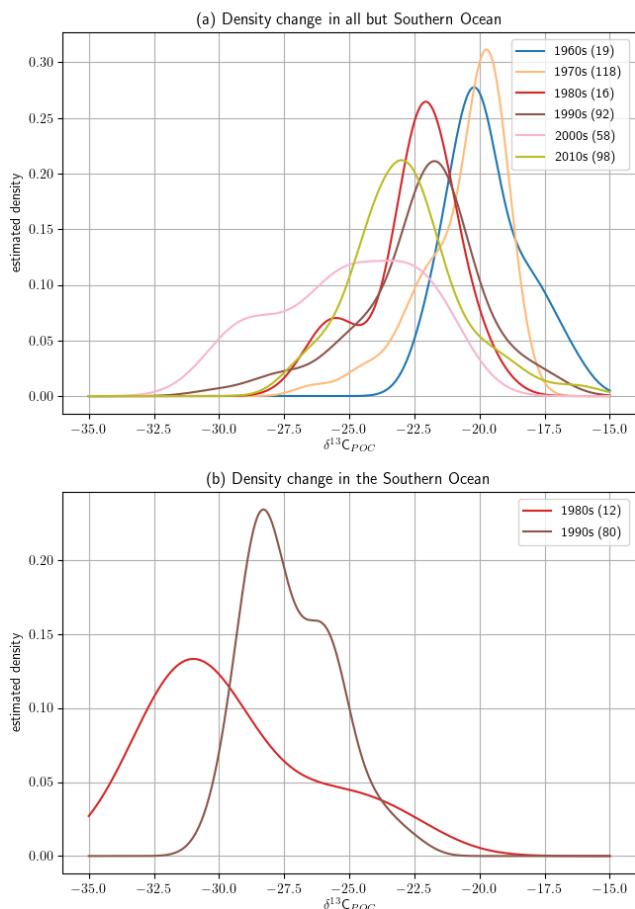


Figure 11. The decadal shift in $\delta^{13}\text{C}_{\text{POC}}$ values for all but the Southern Ocean (a) and only the Southern Ocean (b) shown by estimated densities of $\delta^{13}\text{C}_{\text{POC}}$ values. The differently colored graphs refer to the individual decades. Southern Ocean data are sparsely covered, and the region does not provide enough data for a reasonable comparison.

Median change of $\delta^{13}\text{C}_{\text{POC}}$ in the euphotic zone in all but Southern Ocean

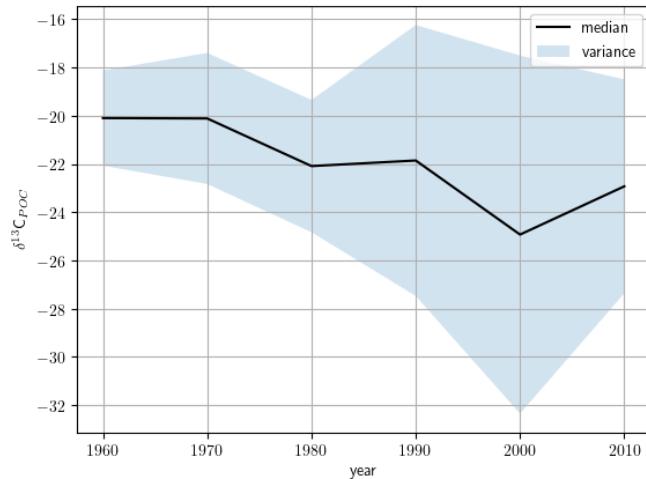


Figure 12. The decadal shift in $\delta^{13}\text{C}_{\text{POC}}$ values in the uppermost 130 m for all but the Southern Ocean: $\delta^{13}\text{C}_{\text{POC}}$ decadal median against the decades. The shaded area around the graph marks the variance of the respective decade in each direction.

show our $\delta^{13}\text{C}_{\text{POC}}$ data values are mostly located between $\delta^{13}\text{C}_{\text{POC}} = -15\text{\textperthousand}$ and $\delta^{13}\text{C}_{\text{POC}} = -35\text{\textperthousand}$ with two maxima at around $\delta^{13}\text{C}_{\text{POC}} = -27\text{\textperthousand}$ and $\delta^{13}\text{C}_{\text{POC}} = -23\text{\textperthousand}$, the latter one being more pronounced. A comparison of the main sample methods showed consistent results when compared with regions. $\delta^{13}\text{C}_{\text{POC}}$ data separated by months indicate counteracting seasonal trends in both hemispheres, but more data are required to demonstrate robust seasonality.

The interpolated data provide insights into geographical behavior of the sampled $\delta^{13}\text{C}_{\text{POC}}$ data. We showed a good general coverage of all global oceans by $\delta^{13}\text{C}_{\text{POC}}$ but observed a lack of data in PANGAEA that cover North Pacific regions. Since the Atlantic Ocean provides the best coverage, corresponding data were used for a north–south trend anal-

ysis, where we observed that the lowest values ($\lesssim -28\text{\textperthousand}$) can be found in the Southern Ocean, whereas the highest ($\gtrsim -22\text{\textperthousand}$) are restricted to low-latitude regions. This might also have influenced the observed lower $\delta^{13}\text{C}_{\text{POC}}$ values in the Southern Hemisphere compared to the Northern Hemisphere, due to the relatively good coverage of the Southern Ocean. Finally, we showed the sample locations and value development of $\delta^{13}\text{C}_{\text{POC}}$ over the observed decades. Since the Southern Ocean data were mainly sampled in the 1990s, a significant multi-decadal trend could not be detected there. In all other oceans our $\delta^{13}\text{C}_{\text{POC}}$ data show a decrease by about 3‰ over the observed time frame, which is about double the rate of the known Suess effect (Keeling, 1979) on aqueous $\delta^{13}\text{CO}_2$ (Young et al., 2013). This corroborates an increase in phytoplankton carbon fractionation that may be associated with a change in phytoplankton communities as previously suggested (Lorrain et al., 2020; Young et al., 2013). The data set shows promise for better understanding, constraining and prediction of carbon cycling as it provides a validation tool for mechanistic models and supports separation of non-spatial components in $\delta^{13}\text{C}_{\text{POC}}$ variations.

Appendix A: Statistical properties of $\delta^{13}\text{C}_{\text{POC}}$ kernel density estimates

In Tables A1, A2, A3, A4 and A5 we present the modes, medians and confidence limits of the KDEs derived in Figs. 1, 3, 7, 9 and 11, respectively.

Table A1. Statistical properties of the KDE derived for Fig. 1 evaluated on an equidistant grid over $[-55.15, -4.5]$ with 1001 grid points: the first column indicates the respective KDE, the following two its modes, the fourth the median and the fifth the 95 % confidence interval of the respective KDE. All values are given in per mill.

$\delta^{13}\text{C}_{\text{POC}}$ KDE	Dominant mode	Second mode	Median	95 % confidence interval
Figure 1	-23.6	-26.9	-23.8	$[-30.9, -17.0]$

Table A2. Statistical properties of the KDEs derived for Fig. 3 evaluated on an equidistant grid over $[-35, -15]$ with 1001 grid points: the first column indicates the respective KDE, the following two its modes, the fourth the median and the fifth the 95 % confidence interval of the respective KDE. All values are given in per mill.

$\delta^{13}\text{C}_{\text{POC}}$ KDE	Dominant mode	Second mode	Median	95 % confidence interval
Figure 3a, full	-21.8	-24.3	-24.3	$[-26.8, -18.3]$
Figure 3a, bottle	-25.1	–	-24.8	$[-26.9, -22.0]$
Figure 3a, intake	-21.6	–	-20.7	$[-24.0, -17.4]$
Figure 3a, net	-21.6	-27.4	-21.7	$[-26.4, -19.5]$
Figure 3a, trap	-24.3	-21.6	-24.1	$[-27.2, -20.0]$
Figure 3b, full	-21.6	–	-21.3	$[-24.2, -18.0]$
Figure 3b, intake	-19.5	-21.1	-20.3	$[-24.8, -17.2]$
Figure 3b, net	-21.5	–	-21.6	$[-23.9, -19.4]$
Figure 3c, full	-21.6	–	-21.6	$[-26.4, -17.6]$
Figure 3c, bottle	-24.7	–	-24.9	$[-29.8, -21.1]$
Figure 3c, net	-21.6	–	-21.6	$[-24.0, -19.5]$
Figure 3c, trap	-21.8	-18.8	-21.7	$[-22.8, -18.5]$
Figure 3d, full	-24.9	–	-24.6	$[-27.1, -21.9]$
Figure 3d, bottle	-25.2	–	-24.8	$[-26.5, -22.1]$
Figure 3d, net	-22.8	-26.9	-24.2	$[-29.4, -19.7]$
Figure 3d, trap	-24.3	-26.6	-24.5	$[-27.2, -22.9]$

Table A3. Statistical properties of the KDEs derived for Fig. 7 evaluated on an equidistant grid over $[-35, -15]$ with 1001 grid points: the first column indicates the respective KDE, the following two its modes, the fourth the median and the fifth the 95 % confidence interval of the respective KDE. All values are given in per mill.

$\delta^{13}\text{C}_{\text{POC}}$ KDE	Dominant mode	Second mode	Median	95 % confidence interval
Figure 7a, all	-21.8	–	-22.8	$[-29.9, -18.1]$
Figure 7a, biome 9	-24.0	–	-23.8	$[-27.5, -18.5]$
Figure 7a, biome 10	-21.7	–	-21.5	$[-25.0, -17.9]$
Figure 7a, biome 11	-21.6	-21.1	-21.3	$[-24.4, -17.7]$
Figure 7a, biome 12	-21.7	–	-21.9	$[-23.2, -20.8]$
Figure 7a, biome 13	-21.9	-24.9	-22.0	$[-24.4, -20.4]$
Figure 7a, biome 15	-22.7	–	-22.8	$[-26.5, -19.2]$
Figure 7a, biome 16	-28.7	-26.0	-27.7	$[-30.7, -24.1]$
Figure 7a, biome 17	-27.8	–	-28.5	$[-32.7, -24.9]$

Table A4. Statistical properties of the KDEs derived for Fig. 9 evaluated on an equidistant grid over $[-35, -15]$ with 1001 grid points: the first column indicates the respective KDE, the following two its modes, the fourth the median and the fifth the 95 % confidence interval of the respective KDE. All values are given in per mill.

$\delta^{13}\text{C}_{\text{POC}}$ KDE	Dominant mode	Second mode	Median	95 % confidence interval
Figure 9a, Feb	-24.7	–	-25.0	[-29.2, -22.3]
Figure 9a, Mar	-24.3	–	-24.0	[-26.6, -20.3]
Figure 9a, Apr	-19.0	–	-20.4	[-26.2, -16.5]
Figure 9a, May	-24.1	-27.0	-24.0	[-28.3, -19.0]
Figure 9a, Jun	-24.0	-27.9	-25.2	[-30.5, -20.1]
Figure 9a, Aug	-23.9	–	-23.6	[-27.7, -18.9]
Figure 9a, Sep	-22.0	-26.0	-23.1	[-28.9, -18.7]
Figure 9a, Oct	-21.5	–	-21.5	[-23.4, -19.7]
Figure 9b, Jan	-27.2	–	-26.5	[-28.9, -22.1]
Figure 9b, Feb	-29.9	–	-26.2	[-30.3, -19.8]
Figure 9b, Mar	-23.3	-28.5	-23.3	[-29.0, -21.0]
Figure 9b, Apr	-28.4	-21.6	-28.1	[-32.6, -19.9]
Figure 9b, Sep	-28.9	-20.4	-28.5	[-30.8, -23.6]
Figure 9b, Oct	-28.5	-22.3	-27.7	[-31.7, -20.8]
Figure 9b, Nov	-28.1	–	-27.7	[-31.8, -20.1]
Figure 9b, Dec	-26.7	-24.4	-26.0	[-28.3, -23.3]

Table A5. Statistical properties of the KDEs derived for Fig. 11 evaluated on an equidistant grid over $[-35, -15]$ with 1001 grid points: the first column indicates the respective KDE, the following two its modes, the fourth the median and the fifth the 95 % confidence interval of the respective KDE. All values are given in per mill.

$\delta^{13}\text{C}_{\text{POC}}$ KDE	Dominant mode	Second mode	Median	95 % confidence interval
Figure 11a, 1960s	-20.0	–	-19.9	[-26.8, -16.5]
Figure 11a, 1970s	-19.8	–	-20.4	[-25.0, -18.0]
Figure 11a, 1980s	-21.7	-25.3	-22.1	[-26.9, -18.5]
Figure 11a, 1990s	-21.8	-27.3	-22.1	[-27.6, -18.2]
Figure 11a, 2000s	-22.4	–	-23.2	[-30.4, -19.2]
Figure 11a, 2010s	-23.1	–	-23.3	[-27.4, -17.6]
Figure 11b, 1960s	-27.5	-30.3	-27.7	[-31.4, -25.2]
Figure 11b, 1980s	-31.0	–	-29.8	[-34.3, -15.0]

Author contributions. MTV collected and merged the data, performed the analyses, and drafted the manuscript. CJS initiated and supported the data collection, conducted the grid interpolations, guided analyses of the data, and structured and proofread the manuscript. MS supported the data collection, guided data analyses and proofread the manuscript. RET provided additional data and ideas for data analyses and proofread the manuscript. AL provided additional data and proofread the manuscript. AO guided the analysis of the data and proofread the manuscript. TS guided the elaboration of the manuscript, and structured and proofread it.

Competing interests. The contact author has declared that neither they nor their co-authors have any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. We would like to thank Tronje Kemena for providing the basic global biomes masks, used for analyzing the interpolated data sets on the coarse grid.

We thank the referees and editor for their constructive feedback regarding the initial version of the manuscript.

Financial support. This research has been supported by the Helmholtz School for Marine Data Science (MarDATA) (grant no. HIDSS-0005) and by the Deutsche Forschungsgemeinschaft (DFG) (project no. 445549720).

Review statement. This paper was edited by Attila Demény and reviewed by Anne Morée and one anonymous referee.

References

- AESOPS: U.S. JGOFS Antarctic Environment and Southern Ocean Process Study, available at: <http://usjgofs.whoi.edu/southern.html>, last access: 3 December 2020.
- Alfred-Wegener-Institut: PANGAEA Data Publisher for Earth & Environmental Science, available at: <https://www.pangaea.de>, last access: 3 December 2020.
- Altabet, M. A. and Francois, R.: Natural nitrogen and carbon stable isotopic composition in surface water at cruise NBP96-05, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.128266>, 2003a.
- Altabet, M. A. and Francois, R.: Natural nitrogen and carbon stable isotopic composition of station NBP96-05-06-4, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.128229>, 2003b.
- Banse, K.: New views on the degradation and disposition of organic particles as collected by sediment traps in the open sea, Deep-Sea Res., 37, 1177–1195, [https://doi.org/10.1016/0198-0149\(90\)90058-4](https://doi.org/10.1016/0198-0149(90)90058-4), 1990.
- Bidigare, R. R., Fluegge, A., Freeman, K. H., Hanson, K. L., Hayes, J. M., Hollander, D., Jasper, J. P., King, L. L., Laws, E. A., Milder, J., Millero, F. J., Pancost, R., Popp, B. N., Steinberg, P. A., and Wakeham, S. G.: Consistent fractionation of ^{13}C in nature and in the laboratory: Growth-rate effects in some haptophyte algae, Global Biogeochem. Cy., 11, 279–292, <https://doi.org/10.1029/96gb03939>, 1997.
- Buchanan, P. J., Matear, R. J., Chase, Z., Phipps, S. J., and Bindoff, N. L.: Ocean carbon and nitrogen isotopes in CSIRO Mk3L-COAL version 1.0: a tool for palaeoceanographic research, Geosci. Model Dev., 12, 1491–1523, <https://doi.org/10.5194/gmd-12-1491-2019>, 2019.
- Calvert, S. E.: Stable isotope data of sediment trap P84-4, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.68555>, 2002.
- Calvert, S. E. and Soon, M.: Carbon and nitrogen data measured on water samples from the multiple unit large volume filtration system (MULVFS) during John P. Tully cruise IOS_96-09, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.68556>, 2002.

- GAEA [data set], <https://doi.org/10.1594/PANGAEA.808319>, 2013a.
- Calvert, S. E. and Soon, M.: Carbon and nitrogen data measured on water samples from the multiple unit large volume filtration system (MULVFS) during John P. Tully cruise IOS_96-18, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.808320>, 2013b.
- Calvert, S. E. and Soon, M.: Carbon and nitrogen data measured on water samples from the multiple unit large volume filtration system (MULVFS) during John P. Tully cruise IOS_97-02, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.808321>, 2013c.
- Cassar, N., Laws, E. A., and Popp, B. N.: Carbon isotopic fractionation by the marine diatom *Phaeodactylum tricornutum* under nutrient- and light-limited growth conditions, *Geochim. Cosmochim. Ac.*, 70, 5323–5335, <https://doi.org/10.1016/j.gca.2006.08.024>, 2006.
- Chang, A. S., Bertram, M. A., Ivanochko, T. S., Calvert, S. E., Dallimore, A., and Thomson, R. E.: (Supplement 2) Total mass flux, geochemistry and abundance of selected diatom taxa of Effingham Inlet OSU Trap samples, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.806329>, 2013.
- Close, H. G. and Henderson, L. C.: Open-Ocean Minima in $\delta^{13}\text{C}$ Values of Particulate Organic Carbon in the Lower Euphotic Zone, *Frontiers in Marine Science*, 7, 540165, <https://doi.org/10.3389/fmars.2020.540165>, 2020.
- Degens, E. T., Behrendt, M., Gotthardt, B., and Reppmann, E.: Metabolic fractionation of carbon isotopes in marine plankton – II. Data on samples collected off the coasts of Peru and Ecuador, *Deep Sea Research and Oceanographic Abstracts*, 15, 11–20, [https://doi.org/10.1016/0011-7471\(68\)90025-9](https://doi.org/10.1016/0011-7471(68)90025-9), 1968.
- De Jonge, C., Stadnitskaia, A., Hopmans, E. C., Cherkashov, G. A., Fedotov, A., Streletskaia, I., Vasiliev, A. A., and Sinninghe Damsté, J. S.: (Table 2) Particulate organic carbon content and the stable carbon isotope signal of suspended particulate matter samples, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.877962>, 2015a.
- De Jonge, C., Stadnitskaia, A., Hopmans, E. C., Cherkashov, G. A., Fedotov, A., Streletskaia, I., Vasiliev, A. A., and Sinninghe Damsté, J. S.: Drastic changes in the distribution of branched tetraether lipids in suspended matter and sediments from the Yenisei River and Kara Sea (Siberia): Implications for the use of brGDGT-based proxies in coastal marine sediments., *Geochim. Cosmochim. Ac.*, 165, 200–225, <https://doi.org/10.1016/j.gca.2015.05.044>, 2015b.
- Eadie, B. J. and Jeffrey, L. M.: $\delta^{13}\text{C}$ analyses of oceanic particulate matter, *Mar. Chem.*, 1, 199–209, [https://doi.org/10.1016/0304-4203\(73\)90004-2](https://doi.org/10.1016/0304-4203(73)90004-2), 1973.
- Eide, M., Olsen, A., Ninnemann, U. S., and Johannessen, T.: A global ocean climatology of preindustrial and modern ocean $\delta^{13}\text{C}$, *Global Biogeochem. Cy.*, 31, 515–534, <https://doi.org/10.1002/2016gb005473>, 2017.
- EurOBIS Data Management Team: PANGAEA – data from Archive of Ocean Data, available at: http://ipt.vliz.be/eurobis/resource?r=pangaea_2724, last access: 3 December 2020.
- Fay, A. R. and McKinley, G. A.: Global open-ocean biomes: mean and temporal variability, *Earth Syst. Sci. Data*, 6, 273–284, <https://doi.org/10.5194/essd-6-273-2014>, 2014.
- Fischer, G.: Stabile Kohlenstoff-Isotopen in partikulärer organischer Substanz aus dem Südpolarmeer (Atlantischer Sektor), PhD thesis, Bremen University, Bremen, Germany, 1989.
- Fontugne, M. and Duplessy, J. C.: Carbon isotope ratios of marine plankton related to surface water masses, *Earth Planet. Sc. Lett.*, 41, 365–371, [https://doi.org/10.1016/0012-821X\(78\)90191-7](https://doi.org/10.1016/0012-821X(78)90191-7), 1978.
- Fontugne, M. and Duplessy, J. C.: Oceanic carbon isotopic fractionation by marine plankton in the temperature range of -1 to 31°C , *Oceanol. Acta*, 4, 85–90, 1981.
- Fontugne, M., Descolas-Gros, C., and de Billy, G.: The dynamics of CO_2 fixation in the Southern Ocean as indicated by carboxylase activities and organic carbon isotopic ratios, *Mar. Chem.*, 35, 371–380, [https://doi.org/10.1016/S0304-4203\(09\)90029-9](https://doi.org/10.1016/S0304-4203(09)90029-9), 1991.
- Francois, R., Atlabet, M. A., Goericke, R., McCorkle, D. C., Brunet, C., and Posson, A.: Changes in the $\delta^{13}\text{C}$ of surface water particulate organic matter across the subtropical convergence in the SW Indian Ocean, *Global Biogeochem. Cy.*, 7, 627–644, <https://doi.org/10.1029/93GB01277>, 1993.
- Freeman, K. H. and Hayes, J. M.: Fractionation of carbon isotopes by phytoplankton and estimates of ancient CO_2 levels, *Global Biogeochem. Cy.*, 6, 185–198, <https://doi.org/10.1029/92GB00190>, 1992.
- Fry, B.: $^{13}\text{C}/^{12}\text{C}$ fractionation by marine diatoms, *Mar. Ecol. Prog. Ser.*, 134, 283–294, <https://doi.org/10.3354/meps134283>, 1996.
- Fry, B. and Sherr, E. B.: $\delta^{13}\text{C}$ Measurements as Indicators of Carbon Flow in Marine and Freshwater Ecosystems, in: *Stable Isotopes in Ecological Research. Ecological Studies (Analysis and Synthesis)*, edited by: Rundel, P. W., Ehleringer, J. R., and Nagy, K. A., Springer, New York, NY, USA, vol. 68, 196–229, https://doi.org/10.1007/978-1-4612-3498-2_12, 1989.
- Garcia, H. E., Weathers, K., Paver, C. R., Smolyar, I., Boyer, T. P., Locarnini, R. A., Zweng, M. M., Mishonov, A. V., Baranova, O. K., Seidov, D., and Reagan, J. R.: Dissolved Inorganic Nutrients (phosphate, nitrate and nitrate+nitrite, silicate), *World Ocean Atlas 2018*, 4, 35 pp., nOAA ATLAS NESDIS 84, NOAA National Centers for Environmental Information (NCEI), Silver Spring, Maryland, USA, 2018.
- Goericke, R.: Variations of marine plankton $\delta^{13}\text{C}$ with latitude, temperature, and dissolved CO_2 in the world ocean, *Global Biogeochem. Cy.*, 8, 85–90, <https://doi.org/10.1029/93GB03272>, 1994.
- Gruber, N., Keeling, C. D., Bacastow, R. B., Guenther, P. R., Lueker, T. J., Wahlen, M., Meijer, H. A. J., Mook, W. G., and Stocker, T. F.: Spatiotemporal patterns of carbon-13 in the global surface oceans and the oceanic suess effect, *Global Biogeochem. Cy.*, 13, 307–335, <https://doi.org/10.1029/1999GB900019>, 1999.
- Hayes, J. M.: An Introduction to Isotopic Calculations, Woods Hole Oceanographic Institution, available at: http://www.whoi.edu/cms/files/jhayes/2005/9/IsoCalcs30Sept04_5183.pdf (last access: 12 May 2020), 2004.
- Hofmann, M., Wolf-Gladrow, D. A., Takahashi, T., Sutherland, S. C., Six, K. D., and Maier-Reimer, E.: Stable carbon isotope distribution of particulate organic matter in the ocean: a model study, *Mar. Chem.*, 72, 131–150, [https://doi.org/10.1016/s0304-4203\(00\)00078-5](https://doi.org/10.1016/s0304-4203(00)00078-5), 2000.

- IPCC: Summary for policymakers, Cambridge University Press, Cambridge, UK, 3–29, <https://doi.org/10.1017/CBO9781107415324.004>, 2013.
- IPCC: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Core Writing Team, Pachauri, R. K., and Meyer, L. A., Geneva, Switzerland, 2014.
- Jahn, A., Lindsay, K., Giraud, X., Gruber, N., Otto-Bliesner, B. L., Liu, Z., and Brady, E. C.: Carbon isotopes in the ocean model of the Community Earth System Model (CESM1), *Geosci. Model Dev.*, 8, 2419–2434, <https://doi.org/10.5194/gmd-8-2419-2015>, 2015.
- Jasper, J. P. and Hayes, J. M.: A carbonisotopic record of CO₂ levels during the late Quaternary, *Nature*, 347, 462–464, <https://doi.org/10.1038/347462a0>, 1990.
- JGOFS: Joint Global Ocean Flux Study, available at: <http://ijgofs.whoi.edu>, last access: 3 December 2020.
- Kaiser, D., Konovalov, S. K., Arz, H. W., Voss, M., Krüger, S., Pollehnne, F., Jeschek, J., and Waniek, J. J.: Black Sea water column dissolved nutrients and dissolved and particulate organic matter from winter 2013, Maria S. Merian cruise MSM33, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.898717>, 2019.
- Keeling, C. D.: The Suess effect: 13Carbon-14Carbon interrelations, *Environ. Int.*, 2, 229–300, [https://doi.org/10.1016/0160-4120\(79\)90005-9](https://doi.org/10.1016/0160-4120(79)90005-9), 1979.
- Kessler, W. S. and McCreary, J. P.: The annual wind-driven Rossby wave in the subthermocline equatorial Pacific, *J. Phys. Oceanogr.*, 23, 1192–1207, 1992.
- Laws, E. A., Popp, B. N., Bidigare, R. R., Kennicutt, M. C., and Macko, S. A.: Dependence of phytoplankton carbon isotopic composition on growth rate and [CO₂]aq: Theoretical considerations and experimental results, *Geochim. Cosmochim. Ac.*, 59, 1131–1138, [https://doi.org/10.1016/0016-7037\(95\)00030-4](https://doi.org/10.1016/0016-7037(95)00030-4), 1995.
- Lein, A. Y. and Ivanov, M. V.: (Table 9.4.3) Concentrations of suspended matter in water samples from the 9°50' N EPR hydrothermal field and contents and isotopic compositions of organic carbon in suspended matter, PANGAEA [data set], PANGAEA, <https://doi.org/10.1594/PANGAEA.771566>, 2009.
- Lein, A. Y., Bogdanov, Y. A., Grichuk, D. V., Rusanov, I. I., and Sagalevich, A. M.: (Table 5) Concentration of particulate organic carbon and its isotopic composition in water samples from hydrothermal fields at the axis of the East Pacific Rise near 9°50' N, PANGAEA [data set], PANGAEA, <https://doi.org/10.1594/PANGAEA.745910>, 2006.
- Lein, A. Y., Bogdanova, O. Y., Bogdanov, Y. A., and Magazina, L. O.: (Table 6) Isotopic composition of organic carbon from microbial communities within the Lost City hydrothermal field, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.765164>, 2007.
- Levin, I., Schuchard, J., Kromer, B., and Münnich, K. O.: The Continental European Suess Effect, *Radiocarbon*, 31, 431–440, <https://doi.org/10.1017/s0033822200012017>, 1989.
- Liu, B., Six, K. D., and Ilyina, T.: Incorporating the stable carbon isotope ¹³C in the ocean biogeochemical component of the Max Planck Institute Earth System Model, *Biogeosciences*, 18, 4389–4429, <https://doi.org/10.5194/bg-18-4389-2021>, 2021.
- Lorrain, A., Pethybridge, H., Cassar, N., Receveur, A., Allain, V., Bodin, N., Bopp, L., Choy, C. A., Duffy, L., Fry, B., Goni, N., Graham, B. S., Hobday, A. J., Logan, J. M., Ménard, F., Menkes, C. E., Olson, R. J., Pagendam, D. E., Point, D., Revill, A. T., Somes, C. J., and Young, J. W.: Trends in tuna carbon isotopes suggest global changes in pelagic phytoplankton communities, *Glob. Change Biol.*, 26, 458–470, <https://doi.org/10.1111/gcb.14858>, 2020.
- MacKenzie, K. M., Robertson, D. R., Adams, J. N., Altieri, A. H., and Turner, B. L.: Carbon and nitrogen stable isotope data from organisms in the Bay of Panama ecosystem, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.903842>, 2019.
- Magozzi, S., Yool, A., Zandee, H. B. V., Wunder, M. B., and Trueman, C. N.: Using ocean models to predict spatial and temporal variation in marine carbon isotopes, *Ecosphere*, 8, e01763, <https://doi.org/10.1002/ecs2.1763>, 2017.
- McConaughey, T. and McRoy, C. P.: Food-Web structure and the fractionation of Carbon isotopes in the bering sea, *Mar. Biol.*, 53, 257–262, <https://doi.org/10.1007/bf00952434>, 1979.
- Morée, A. L., Schwinger, J., and Heinze, C.: Southern Ocean controls of the vertical marine $\delta^{13}\text{C}$ gradient – a modelling study, *Biogeosciences*, 15, 7205–7223, <https://doi.org/10.5194/bg-15-7205-2018>, 2018.
- Ndeye, M., Sène, M., Diop, D., and Saliège, J.-F.: Anthropogenic CO₂ in the Dakar (Senegal) Urban Area Deduced from ¹⁴C Concentration in Tree Leaves, *Radiocarbon*, 59, 1009–1019, <https://doi.org/10.1017/rdc.2017.48>, 2017.
- NOAA's Pacific Marine Environmental Laboratory: Ferret Support, available at: <http://ferret.pmel.noaa.gov/Ferret>, last access: 26 November 2020.
- Popp, B. N., Takigiku, R., Hayes, J. M., Louda, J. W., and Baker, E. W.: The post-palaeozoic chronology and mechanism of ¹³C depletion in primary marine organic matter, *Am. J. Sci.*, 289, 436–454, 1989.
- Popp, B. N., Laws, E. A., Bidigare, R. R., Dore, J. E., Hanson, K. L., and Wakeham, S. G.: Effect of Phytoplankton Cell Geometry on Carbon Isotopic Fractionation, *Geochim. Cosmochim. Ac.*, 62, 69–77, [https://doi.org/10.1016/s0016-7037\(97\)00333-5](https://doi.org/10.1016/s0016-7037(97)00333-5), 1998.
- Rau, G. H., Takahashi, T., and Des Marais, D. J.: Latitudinal variations in plankton $\delta^{13}\text{C}$: implications for CO₂ and productivity in past oceans, *Nature*, 341, 516–518, <https://doi.org/10.1038/341516a0>, 1989.
- Rau, G. H., Riebesell, U., and Wolf-Gladrow, D.: A model of photosynthetic ¹³C fractionation by marine phytoplankton based on diffusive molecular CO₂ uptake, *Mar. Ecol. Prog. Ser.*, 133, 275–285, 1996.
- Rounick, J. S. and Winterbourn, M. J.: Stable carbon isotopes and carbon flow in ecosystems – Measuring ¹³C to ¹²C ratios can help to trace carbon pathways, *BioScience*, 36, 171–177, <https://doi.org/10.2307/1310304>, 1986.
- Rubino, M., Etheridge, D. M., Trudinger, C. M., Allison, C. E., Battle, M. O., Langenfelds, R. L., Steele, L. P., Curran, M., Bender, M., White, J. W. C., Jenk, T. M., Blunier, T., and Francey, R. J.: A revised 1000 year atmospheric $\delta^{13}\text{C}$ -CO₂ record from Law Dome and South Pole, Antarctica, *J. Geophys. Res.-Atmos.*, 118, 8482–8499, <https://doi.org/10.1002/jgrd.50668>, 2013.
- Sacket, W. M., Eckelmann, W. R., Bender, M. L., and Bé, A. W. H.: Temperature Dependence of Carbon Isotope Composi-

- tion in Marine Plankton and Sediments, *Science*, 148, 235–237, <https://doi.org/10.1126/science.148.3667.235>, 1965.
- Saupe, S. M., Schell, D. M., and Griffiths, W. B.: Carbon-isotope ratio gradients in western arctic zooplankton, *Mar. Biol.*, 103, 427–432, <https://doi.org/10.1007/BF00399574>, 1989.
- Schmittner, A. and Somes, C. J.: Complementary constraints from carbon (^{13}C) and nitrogen (^{15}N) isotopes on the glacial ocean's soft-tissue biological pump, *Paleoceanography*, 31, 669–693, <https://doi.org/10.1002/2015PA002905>, 2016.
- Schmittner, A., Gruber, N., Mix, A. C., Key, R. M., Tagliabue, A., and Westberry, T. K.: Biology and air-sea gas exchange controls on the distribution of carbon isotope ratios ($\delta^{13}\text{C}$) in the ocean, *Biogeosciences*, 10, 5793–5816, <https://doi.org/10.5194/bg-10-5793-2013>, 2013.
- Silverman, B. W.: Density Estimation for Statistics and Data Analysis, Monographs on Statistics and Applied Probability, Chapman and Hall, London, UK, 1986.
- Suess, E.: Particulate organic carbon flux in the oceans—surface productivity and oxygen utilization, *Nature*, 288, 260–263, 1980.
- Tagliabue, A. and Bopp, L.: Towards understanding global variability in ocean carbon-13, *Global Biogeochem. Cy.*, 22, GB1025, <https://doi.org/10.1029/2007gb003037>, 2008.
- Thiede, J., Gerlach, S. A., Altenbach, A., and Henrich, R.: Sedimentation im europaeischen Nordmeer – Organisation und Forschungsprogramm des Sonderforschungsbereiches 313 fuer den Zeitraum 1988–1990, Tech. rep., Kiel University, Kiel, Germany, 1988.
- Tjiputra, J. F., Schwinger, J., Bentsen, M., Morée, A. L., Gao, S., Bethke, I., Heinze, C., Goris, N., Gupta, A., He, Y.-C., Olivie, D., Seland, Ø., and Schulz, M.: Ocean biogeochemistry in the Norwegian Earth System Model version 2 (NorESM2), *Geosci. Model Dev.*, 13, 2393–2431, <https://doi.org/10.5194/gmd-13-2393-2020>, 2020.
- Trull, T. W. and Armand, L. K.: Insights into Southern Ocean carbon export from the $\delta^{13}\text{C}$ of particles and dissolved inorganic carbon during the SOIREE iron release experiment, *Deep-Sea Res. Pt. II*, 48, 2655–2680, [https://doi.org/10.1016/S0967-0645\(01\)00013-3](https://doi.org/10.1016/S0967-0645(01)00013-3), 2001.
- Trull, T. W. and Armand, L. K.: $\delta^{13}\text{C}$ content of particulate organic carbon measured on samples from traps during TANGAROA cruise SOIREE, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.807904>, 2013a.
- Trull, T. W. and Armand, L. K.: $\delta^{13}\text{C}$ content of fractionated particulate organic carbon measured on samples from traps during TANGAROA cruise SOIREE, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.807906>, 2013b.
- Tuerena, R. E., Ganeshram, R. S., Humphreys, M. P., Browning, T. J., Bouman, H., and Piotrowski, A. P.: Isotopic fractionation of carbon during uptake by phytoplankton across the South Atlantic subtropical convergence, *Biogeosciences*, 16, 3621–3635, <https://doi.org/10.5194/bg-16-3621-2019>, 2019.
- Verwega, M.-T., Somes, C. J., Tuerena, R. E., and Lorrain, A.: A global marine particulate organic carbon-13 isotope data product, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.929931>, 2021.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, *Nat. Methods*, 17, 261–272, <https://doi.org/10.1038/s41592-019-0686-2>, 2020.
- Volk, T. and Hoffert, M. I.: Ocean carbon pumps: analysis of relative strengths and efficiencies in ocean-driven atmospheric CO₂ changes, American Geophysical Union; Geophysical Monograph, 32, 99–110, 1985.
- Voss, M. and von Bodungen, B.: Carbon and nitrogen from mooring NB2, PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.106805>, 2003.
- Wada, E., Terazaki, M., Kabaya, Y., and Nemoto, T.: ^{15}N and ^{13}C abundances in the Antarctic Ocean with emphasis on the biogeochemical structure of the food web, *Deep-Sea Res.*, 34, 829–841, [https://doi.org/10.1016/0198-0149\(87\)90039-2](https://doi.org/10.1016/0198-0149(87)90039-2), 1987.
- Westerhausen, L. and Sarnthein, M.: $\delta^{13}\text{C}$ of plankton from surface water (Table A2), PANGAEA [data set], <https://doi.org/10.1594/PANGAEA.89388>, 2003.
- Young, J. N., Bruggeman, J., Rickaby, R. E. M., Erez, J., and Conte, M.: Evidence for changes in carbon isotopic fractionation by phytoplankton between 1960 and 2010, *Global Biogeochem. Cy.*, 27, 505–515, <https://doi.org/10.1002/gbc.20045>, 2013.
- Zeebe, R. E. and Wolf-Gladrow, D.: CO₂ in Seawater: Equilibrium, Kinetics, Isotopes, Elsevier Science B.V., Elsevier Oceanography Series, Amsterdam, the Netherlands, 65, 2001.
- Zhang, J., Quay, P., and Wilbur, D.: Carbon isotope fractionation during gas-water exchange and dissolution of CO₂, *Geochim. Cosmochim. Ac.*, 59, 107–114, [https://doi.org/10.1016/0016-7037\(95\)91550-d](https://doi.org/10.1016/0016-7037(95)91550-d), 1995.